

Europäisches Patentamt

European Patent Office

Office européen des brevets



1) Publication number:

0 346 710 B1

12

### **EUROPEAN PATENT SPECIFICATION**

45 Date of publication of patent specification: 10.11.93 51 Int. Cl.<sup>5</sup>: C12N 15/12, C12N 5/10,

C12P 21/02

(21) Application number: 89110096.8

② Date of filing: 03.06.89

- (S) cDNAs coding for members of the carcinoembryonic antigen family.
- Priority: 16.06.88 US 207678 21.11.88 US 274107
- ② Date of publication of application: 20.12.89 Bulletin 89/51
- Publication of the grant of the patent: 10.11.93 Bulletin 93/45
- Designated Contracting States:
   AT BE CH DE ES FR GB GR IT LI NL SE
- 69 References cited: EP-A- 263 933 EP-A- 0 212 880

BIOCHEM. BIOPHYS. RES. COMMUN., vol. 142, no. 2, 30th January 1987, pages 511-518; R. OIKAWA et al.: "Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequence"

MOL. CELL. BIOL., vol. 7, 1987, page 3221-3230; R. BEAUCHEMIN et al.: "Isolation and characterization of full-lenght functional cDNA clones for human carcinoembryonic antigen"

- 73 Proprietor: MILES INC.
  One Mellon Center
  500 Grant Str.
  Pittsburgh, PA 15219-2502(US)
- 27 Inventor: Barnett, Thomas R., Dr. 27 Jeffrey Road
  East Haven, CT 06513(US)
  Inventor: Elting, James J., Dr. 5 Heatherwood Drive
  Madison, CT 06443(US)
  Inventor: Kamarck, Michael E. 86 Russell Road
  Bethany, CT 06525(US)
  Inventor: Kretschmer, Axel, Dr. Richard-Zörner-Strasse 32
  D-5060 Bergisch Gladbach 1(DE)
- (24) Representative: Dänner, Klaus, Dr. et al Bayer AG Konzernverwaltung RP Patente Konzern D-51368 Leverkusen (DE)

BEST AVAILABLE COPY

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid (Art. 99(1) European patent convention).

STR004 #13 PROC. NATL. ACAD. SCI. USA, vol. 85, September 1988, pages 6959-6963; Y. HINODA et al.: "Molecular cloning of a cDNA coding biliary glycoprotein I: primary structure of a glycoprotein immunologically crossreactive with carcinoembryonic antigen"

GENE, vol. 71, no. 2, November 1988, pages 439-449; B.C. ROONEY et al.: "Molecular cloning of a cDNA for human pregnancy-specific B1-glycoprotein: homology with human carcinoembryonic antigen and related proteins"

### Description

#### BACKGROUND OF THE INVENTION

#### 5 Field of the Invention

The present invention concerns nucleic acid sequences which code for carcinoembryonic antigen (CEA) antigen family peptide sequences.

#### Background Information

40

45

50

Carcinoembryonic antigen was first described by Gold and Freedman, J. Exp. Med., 121, 439-462, (1965). CEA is characterized as a glycoprotein of approximately 200,000 molecular weight with 50-60% by weight of carbohydrate. CEA is present during normal human fetal development, but only in very low concentration in the normal adult intestinal tract. It is produced and secreted by a number of different tumors.

CEA is a clinically useful tumor marker for the management of colorectal cancer patients. CEA can be measured using sensitive immunoassay methods. When presurgical serum levels of CEA are elevated, a postsurgical drop in serum CEA to the normal range typically indicates successful resection of the tumor. Postsurgical CEA levels that do not return to normal often indicate incomplete resection of the tumor or the presence of additional tumor sites in the patient. After returning to normal levels, subsequent rapid rises in serum CEA levels usually indicate the presence of metastages. Slower postsurgical rises from the normal level are most often interpreted to indicate the presence of new primary tumors not previously detected. Post surgical management of colon cancer patients is thus facilitated by the measurement of CEA.

CEA is a member of an antigen family. Because of this, the immunoassay of CEA by presently available methods is complicated by the fact that CEA is but one of several potentially reactive antigens. There have been at least sixteen CEA-like antigens described in the literature. Since some of these appear to be the same antigen described by different investigators, the actual number of different antigens is somewhat less than this number. Nonetheless, there is a complex array of cross-reactive antigens which can potentially interfere with an immunoassay of the CEA released by tumors. It is known that serum levels of CEA-like antigens are elevated in many non-cancerous conditions such an inflammatory liver diseases and also in smokers. It is important that immunoassays used for the monitoring of cancer patient status not be interfered with by these other CEA-like antigens. Conversely, it is important to be able to distinguish the antigens by immunoassays because of the possibility that different tumor types may preferentially express different forms of CEA. If so, then the ability to reliably measure the different forms of CEA can provide the means to diagnose or more successfully treat different forms of cancer.

The members of the "CEA family" share some antigenic determinants. These common epitopes are not useful in distinguishing the members of the antigen family and antibodies recognizing them are of little use for measuring tumor-specific CEA levels.

- U.S.P. 3,663,684, entitled "Carcinoembryonic Antigen and Diagnostic Method Using Radioactive Iodine", concerns purification and radioiodination of CEA for use in a RIA.
- U.S.P. 3,697,638 describes that CEA is a mixture of antigens (components A and B in this case). U.S.P. 3,697,638 mentions methods for separating and radioiodinating each component and their use in specific RIA's.
- U.S.P. 3,852,415, entitled "Compositions for Use in Radioimmunoassay, as Substitute for Blood Plasma Extract in Determination of Carcinoembryonic Antigen" relates to the use of a buffer containing EDTA and bovine serum albumin as a substitute for plasma as a diluent for CEA RIA's.
- U.S.P. 3,867,363, entitled "Carcinoembryonic Antigens", is directed to the isolation of CEA components A and B, their labelling and use in a RIA.
- U.S.P. 3,927,193, entitled "Localization of Tumors by Radiolabelled Antibodies", concerns the use of radiolabelled anti-CEA antibodies in whole body tumor imaging.
- U.S.P. 3,956,258, entitled "Carcinoembryonic Antigens", relates to the isolation of CEA components A and B.
- U.S.P. 4,086,217, entitled "Carcinoembryonic Antigens", is directed to the isolation of CEA components A and B.
- U.S.P. 4,140,753, entitled "Diagnostic Method and Reagent", concerns the purification of a CEA isomer called CEA-S1 and its use in a RIA.
  - U.S.P. 4,145,336, entitled "Carcinoembryonic Antigen Isomer", relates to the antigen CEA-S1.

- U.S.P. 4,180,499, entitled "Carcinoembryonic Antigens", describes a process for producing CEA component B.
- U.S.P. 4,228,236, entitled "Process of Producing Carcinoembryonic Antigen", is directed to the use of the established cell lines LS-174T and LS-180 or clones or derivatives thereof for the production of CEA.
- U.S.P. 4,272,504, entitled "Antibody Adsorbed Support Method for Carcinoembryonic Antigen Assay", concerns two concepts for the radioimmunoassay of CEA. First, U.S.P. 4,272,504 relates to a sample pretreatment in the form of heating to 65 to 85 °C at pH 5 to precipitate and eliminate extraneous protein. Second, it describes the use of a solid phase antibody (either on beads or tubes) as a means to capture analyte and radiolabelled CEA tracer.
- U.S.P. 4,299,815, entitled "Carcinoembryonic Antigen Determination", concerns diluting a CEA sample with water and pretreating by heating to a temperature below which precipitation of protein will occur. The pretreated sample is then immunoassayed using RIA, EIA, FIA or chemiluminescent immunoassay.
- U.S.P. 4,349,528, entitled "Monoclonal Hybridoma Antibody Specific for High Molecular Weight Carcinoembryonic Antigen", is directed to a monoclonal antibody reacting with 180 kD CEA, but not with other molecular weight forms.
- U.S.P. 4,467,031, entitled "Enzyme-Immunoassay for Carcinoembryonic Antigen", relates to a sandwich enzyme immunoassay for CEA in which the first of two anti-CEA monoclonal antibodies is attached to a solid phase and the second monoclonal is conjugated with peroxidase.
- U.S.P. 4,489,167, entitled "Methods and Compositions for Cancer Detection", describes that CEA shares an antigenic determinant with alpha-acid glycoprotein (AG), which is a normal component of human serum. The method described therein concerns a solid-phase sandwich enzyme immunoassay using as one antibody an antibody recognizing AG and another antibody recognizing CEA, but not AG.
- U.S.P. 4,578,349, entitled "Immunoassay for Carcinoembryonic Antigen (CEA)", is directed to the use of high salt containing buffers as diluents in CEA immunoassays.
- EP 113072-A, entitled "Assaying Blood Sample for Carcinoembryonic Antigen After Removal of Interfering Materials by Incubation with Silica Gel", relates to the removal from a serum of a plasma sample of interfering substances by pretreatment with silica gel. The precleared sample is then subjected to an immunoassay.
- EP 102008-A, entitled "Cancer Diagnostics Carcinoembryonic Antigen Produced from Perchloric Acid Extracts Without Electrophoresis", relates to a procedure for the preparation of CEA from perchloric acid extracts, without the use of an electrophoresis step.
- EP 92223-A, entitled "Determination of Carcinoembryonic Antiyen in Cytosol or Tissue for Therapy Control and Early Recognition of Regression", concerns an immunoassay of CEA, not in serum or plasma, but in the cytosol fraction of the tumor tissue itself.
- EP 83103759.6, entitled "Cytosole-CEA-Measurement as Predictive Test in Carcinoma, Particularly Mammacarcinoma", is similar to EP 92223-A.
- EP 83303759, entitled "Monoclonal Antibodies Specific to Carcinoembryonic Antigen", relates to the production of "CEA specific" monoclonal antibodies and their use in immunoassays.
- WO 84/02983, entitled "Specific CEA-Family Antigens, Antibodies Specific Thereto and Their Methods of Use", is directed to the use of monoclonal antibodies to CEA-meconium (MA)-, and NCA-specific epitopes in immunoassays designed to selectively measure each of these individual components in a sample.
- All of the heretofore CEA assays utilize either monoclonal or polyclonal antibodies which are generated by immunizing animals with the intact antigen of choice. None of them address the idea of making sequence specific antibodies for the detection of a unique primary sequence of the various antigens. They do not cover the use of any primary amino acid sequence for the production of antibodies to synthetic peptides or fragments of the natural product. They do not include the concept of using primary amino acid sequences to distinguish the CEA family members. None of them covers the use of DNA or RNA clones for isolating the genes with which to determine the primary sequence.

50

35

### **DEFINITIONS**

<del></del>		
	A	adenine
	G	guanine

Nucleic Acid Abbreviations

C cytosine
T thymidine

U uracil

### Amino Acid Abbreviations:

	Amino A	Cld ADD	reviations:	
15			Asp	aspartic acid
			Asn	asparagine
			Thr	threonine
20			Ser	serine
20			Glu	glutamic acid
			Gln	glutamine
			Pro	proline
25			Gly	glycine
	, ••		Ala	alanine
			Cys	cysteine
30			Val	valine
		•	Met	methionine
			Ile	isoleucine
35			Leu	leucine
			Tyr	tyrosine
			Phe	phenylalanine
			Trp	tryptophan
40			Lys	lysine
			His	histidine
			Arg	arginine

45

5

10

Nucleotide - A monomeric unit of DNA or RNA containing a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is called a nucleoside. The base characterizes the nucleotide. The four DNA bases are adenine ("A"), guanine ("G"), cytosine ("C"), and thymine ("T"). The four RNA bases are A, G, C and uracil ("U").

DNA Sequence - A linear array of nucleotides connected one to the other by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

Functional equivalents - It is well known in the art that in a DNA sequence some nucleotides can be replaced without having an influence on the sequence of the expression product. With respect to the peptide this term means that one or more amino acids which have no function in a particular use can be deleted or replaced by another one.

Codon - A DNA sequence of three nucleotides (a triplet) which encodes through mRNA an amino acid, a translation start signal or a translation termination signal. For example, the nucleotide triplets TTA, TTG,

CTT, CTC, CTA and CTG encode the amino acid leucine ("Leu"), TAG, TAA and TGA are translation stop signals and ATG is a translation start signal.

Reading Frame - The grouping of codons during translation of mRNA into amino acid sequences. During translation, the proper reading frame must be maintained. For example, the sequence GCTGGTTGTAAG may be translated in three reading frames or phases, each of which affords a different amino acid sequence

GCT GGT TGT AAG - Ala-Gly-Cys-Lys
G CTG GTT GTA AG - Leu-Val-Val
GC TGG TTG TAA G - Trp-Leu-(STOP).

Polypeptide - A linear array of amino acids connected one to the other by peptide bonds between the alpha-amino and carboxy groups of adjacent amino acids.

Genome - The entire DNA of a cell or a virus. It includes inter alia the structural genes coding for the polypeptides of the cell or virus, as well as its operator, promoter and ribosome binding and interaction sequences, including sequences such as the Shine-Dalgarno sequences.

Structural Gene - A DNA sequence which encodes through its template or messenger RNA ("mRNA") a sequence of amino acids characteristic of a specific polypeptide.

Transcription - The process of producing mRNA from a structural gene.

Translation - The process of producing a polypeptide from mRNA.

Expression - The process undergone by a structural gene to produce a polypeptide. It is a combination of transcription and translation.

Plasmid - A non-chromosomal double-stranded DNA sequence comprising an intact "replicon" such that the plasmid is replicated in a host cell. When the plasmid is placed within a unicellular organism, the characteristics of that organism may be changed or transformed as a result of the DNA of the plasmid. For example, a plasmid carrying the gene for tetracycline resistance (Tet<sup>R</sup>) transforms a cell previously sensitive to tetracycline into one which is resistant to it. A cell transformed by a plasmid is called a "transformant".

Phage or Bacteriophage - Bacterial virus, many of which consist of DNA sequences encapsulated in a protein envelope or coat ("capsid protein").

Cloning Vehicle - A plasmid, phage DNA or other DNA sequence which is capable of replicating in a host cell, which is characterized by one or a small number of endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without attendant loss of an essential biological function of the DNA, e.g., replication, production of coat proteins or loss of promoter or binding sites, and which contains a marker suitable for use in the identification of transformed cells, e.g., tetracycline resistance or ampicillin resistance. A cloning vehicle is often called a vector.

Cloning - The process of obtaining a population of organisms or DNA sequences derived from one such organism or sequence by asexual reproduction.

Recombinant DNA Molecule or Hybrid DNA - A molecule consisting of segments of DNA from different genomes which have been joined end-to-end outside of living cells and have the capacity to infect some host cell and be maintained therein.

cDNA Expression Vector - A procaroytic cloning vehicle which also contains sequences of nucleotides that facilitate expression of cDNA sequences in eucaroytic cells. These nucleotides include sequences that function as eucaryotic promoter, alternative splice sites and polyadenylation signals.

Transformation/Transfection - DNA or RNA is introduced into cells in such a way as to allow gene expression. "Infected" referred to herein concerns the introduction of RNA or DNA by a viral vector into the host.

"Injected" referrred to herein concerns the microinjection (use of a small syringe) of DNA into a cell.

CEA antigen family (CEA gene family) - a set of genes (gene family) and their products (antigen family) that share nucleotide sequences homologous to partial cDNA LV-7 (CEA-(a)) and as a result of theses similarities also share a subset of their antigenic epitopes. Examples of the CEA antigen family include CEA (= CEA-(b)), transmembrane CEA (TMCEA) = CEA-(c) and normal crossreacting antigen NCA (= CEA-(d)).

10

25

30

### SUMMARY OF THE INVENTION

The present invention concerns the following DNA sequences designated as TM-2 (CEA-(e)), TM-3 (CEA-(f)), TM-4 (CEA-(g)), KGCEA1 and KGCEA2, which code for CEA antigen family peptide sequences:

ACAGGATTCTACACCCTACAAGTCATAAAGTCAGATCTTGTGAATAGAAGAAGCAACTGGA

ThrGlyPheTyrThrLeuGlnVallleLysSerAspLeuValAsnGluGluAlaThrGly

	1030	1050	1070	
5	ATAGTCACTGATAATGCTC IleValThrAspAsnAlaL	TACCACAAGAAAATGGCCT euProGlnGluAsnGlyLe	CTCACCTGGGGCCATTGCTGGC	;
	1090	1110	1130	
10	ATTGTGATTGGAGTAGTGG IleVallleGlyValValA	CCCTGGTTGCTCTGATAG laLeuValAlaLeuIleA	CAGTAGCCCTGGCATGTTTTCTC laValAlaLeuAlaCysPheLev	3
	1150	1170	1190	
15	CATTTCGGGAAGACCGGCA HisPheGlyLysThcGlyA	GGGCAAGCGACCAGCGTGA rgAlaSerAspGlnArgA:	ATCTCACAGAGCACAAACCCTCA spleuThrGluHisLysProSe	\ [
	1210	1230	1250	
20	GTCTCCAACCACACTCAGG ValSerAsnHisThrGlnA	ACCACTCCAATGACCCAC spHisSerAsnAspProP	CTAACAAGATGAATGAAGTTACT roAsnLysMetAsnGluValThi	r
	1270	1290	1310	
25	TATTCTACCCTGAACTTTC TyrSerThrLeuAsnPheC	AAGCCCAGCAACCCACAC	AACCAACTTCAGCCTCCCCATCC	: :
	1330	1350	1370	
30	CTAACAGCCACAGAAATAA LeuThrAlaThrGluIle		AGCAGTAATGAAACCTGTCCTG ysGln	C
35	1390	1410	1430	
33	TCACTGCAGTGCTGATGT	ATTTCAAGTCTCTCACCCT	CATCACTAGGAGATTCCTTTCC	C
	1450	1470	1490	
40	CTGTAGGGTAGAGGGGTGG	GGAČAGAAACAACTTTCT	CCTACTCTTCCTTAATAGG	C
	1510	1530	1550	
45	ATCTCCAGGCTGCCTGGT	CACTGCCCCTCTCTCAGTG	TCAATAGATGAAAGTACATTGG	G
	1570	1590	1610	
50	AGTCTGTAGGAAACCCAA	CCTTCTTGTCATTGAAATT	TGGCAAAGCTGACTTTGGGAAA	Ġ

	1630	1650	1670	
5	AGGGACCAGAACTTCCCC	TCCCTTCCCCTTTTCCCAAC	CTGGACTTGTTTAAACTT	rocc
·	1690	1710	1730	
	TGTTCAGAGCACTCATTC	CTTCCCACCCCAGTCCTGT	CCTATCACTCTAATTCGGA	TTT
10				
	1750	1770	1790	
	GCCATAGCCTTGAGGTTA	TGTCCTTTTCCATTAAGTAC	ATGTGCCAGGAAACAGCGA	GAG
15	1810	1830	1850	
	AGAGAAAGTAAACGGCAG	TAATGCTTCTCCTATTTCTC	CAAAGCCTTGTGTGAACTA	GCA
20	1870	1890	1910	
	AAGAGAAGAAAATCAAAT	ATATAACCAATAGTGAAATG	CCACAGGTTTGTCCACTGT	CAG
25	1930	1950	1970	
	GGTTGTCTACCTGTAGGA	TCAGGGTCTAAGCACCTTGG	TGCTTAGCTAGAATACCAC	CTA
	1990	2010	2030	
30	ATCCTTCTGGCAAGCCTG	TCTTCAGAGAACCCACTAGA	AGCAACTAGGAAAAATCAC	TTG
	2050	2070	2090	
35	CCAAAATCCAAGGCAATT	CCTGATGGAAAATGCAAAAG		CTT
	2110	2130	2150	
40	TATGGGCTCTGTTCAAGG	CAGTGCTGAGAGGGAGGGGT	TATAGCTTCAGGAGGGAAC	CAG
	2170	2190	2210	
45	CTTCTGATAAACACAATC	TGCTAGGAACTTGGGAAAGG		AGC

	2230	2250	2270	
	GATTATTTAAATTGTTAA	AGAATACACAATTTGGGGTA	TTGGGATTTTTCTCCTTTT	CTC
5	2290	2310	2330	
	TGAGACATTCCACCATTT	TAATTTTTGTAACTGCTTAT	TTATGTGAAAAGGGTTATT	TTT
10	2350	2370	2390	
	ACTTAGCTTAGCTATGTC	AGCCAATCCGATTĞCCTTAG	GTGAAAGAAACCACCGAAA	тсс
15	2410	2430	2450	
	CTCAGGTCCCTTGGTCAG	GAGCCTCTCAAGATTTTTT	TGTCAGAGGCTCCAAATAG	AAA
20	2 4 7 0	2490	2510	
	ATAAGAAAAGGTTTTCTT	CATTCATGGCTAGAGCTAGA	TTTAACTCAGTTTCTAGGC	ACC
	2530	2550	2570	
25	TCAGACCAATCATCAACT	ACCATTCTATTCCATGTTTG	CACCTGTGCATTTTCTGTT	TGC
	2590	2610	2630	
30	CCCCATTCACTTTGTCAC	GAAACCTTGGCCTCTGCTAA	GGTGTATTTGGTCCTTGAG	AAG
	2650	2670	2690	
35	TGGGAGCACCCTACAGGC	SACACTATCACTCATGCTGGT	PGGCATTGTTTAČAGCTAGA	.AAG
	2710	2730	2750	
40	CTGCACTGGTGCTAATGG	CCCTTGGGAAATGGGGCTG?	CGAGGAGGAGGATTATAACT	TAG
	2770	2790	2810	
45	GCCTAGCCTCTTTTAACA	AGCCTCTGAAATTTATCTTT	CTTCTATGGGGTCTATAAA	NTGT
	2830	2850	2870	
50	ATCTTATAATAAAAAGG	AAGGACAGGAGGAAGACAGG	CAAATGTACTTCTCACCCA	STCI
50				

TCTACACAGATGGAATCTCTTTGGGGCTAAGAGAAAGGTTTTATTCTATATTGCTTACCT GATCTCATGTTAGGCCTAAGAGGCTTTCTCCAGGAGGATTAGCTTGGAGTTCTCTATACT CAGGTACCTCTTTCAGGGTTTTCTAACCCTGACACGGACTGTGCATACTTTCCCTCATCC 

# SEQUENCE AND TRANSLATION OF CDNA OF TM-3

5				
	10	30	50	
10	CAGCCGTGCTCGAAGCGTT	CCTGGAGCCCAAGCTCTCCT	CCACAGGTGAAGACAGGGC	SA
	70	90	110	
15		ACCTCTCAGCCCCACTTCAC		
	130	150	170	
20		CACTTCTAACCTTCTGGAAC		
25	190	210	230	
		TCAATGTTGCAGAGGGGAAG heAsnValAlaGluGlyLys		
30	250	270	290	
		TTGGCTACAGCTGGTACAA? heGlyTyrSerTrpTyrLys		
35	310	330	350	
	CGTCAAATTGTAGGATATG ArgGlnIleValGlyTyrA	CAATAGGAACTCAACAAGCT	ACCCCAGGGCCCGCAAACA ThrProGlyProAlaAsnS	GC e r
40	370	390	410	
	GGTCGAGAGACAATATACC GlyArgGluThrIleTyrF	CCAATGCATCCCTGCTGATC roAsnalaSerLeuLeulle	CAGAACGTCACCCAGAATG GlnAsnValThrGlnAsnA	AC Sc
45			· · · · · · · · · · · · · · · · · · ·	•

55

	430	450	470
5		CAAGTCATAAAGTCAGATCTT GlnValileLysSerAspLeu	
	490	510	530
10		GAGCTGCCCAAGCCCTCCATC GluLeuProLysProSerIle	
	550	570	590
15	GTGGAGGACAAGGATGCT ValGluAspLysAspAla	GTGGCCTTCACCTGTGAACCT( ValalaPheThrCysGluPro	GAGACTCAGGACACCAACCTAC
	610	630	650
20		CAGAGCCTCCCGGTCAGTCCC GlnSerLeuProValSerPro	
25	670	690	710
		CTCAGTGTCACAAGGAATGACA LeuSerValThrArgAsnAsp'	
30	730	750	770
	ATACAGAACCCAGTGAGT IleGlnAsnProValSer	GCGAACCGCAGTGACCCAGTC AlaAsnArgSerAspProVal	ACCTTGAATGTCACCTATGGC
35	790	810	830
	CCGGACACCCCCACCATT ProAspThrProThrIle	TCCCCTTCAGACACCTATTACC SerProSerAspThrTyrTyr/	CGTCCAGGGGCAAACCTCAGC ArgProGlyAlaAsnLeuSer
40		•	
45			

	850	870	890
5			CTCCTGGCTTATCAATGGAACA rSerTrpLeuIleAsnGlyThr
	910	930	950
10			CACTGTGAATAATAGTGGATCC eThrValAsnAsnSerGlySer
	970	990	1010
15	TATACCTGCCACGCCAATA TyrThrCysHisAlaAsnA	ACTCAGTCACTGGCTGCAA snSerValThrGlyCysAs	CAGGACCACAGTCAAGACGATC
	1030	1050	1070
20			AATCAAAGCCAGCAAGACCACA nlleLysalaSerLysThrThr
	1090	1110	1130
25			CACAAATGACACTGGAATCTCC
	1150	1170	1190
30			GGAGAGGATGAAGCTGTCCCAG rGluArgMetLysLeuSerGln
35	1210	1230	1250
			GGATGCTGGGACGTATTGGTGT UASPAlaGlyThrTyrTrpCys
40			
45			

GAGGTCTTCAACCCAATCAGTAAGAACCAAAGCGACCCCATCATGCTGAACGTAAGCIVAlPheAsnProIleSerLysAsnGlnSerAspProIleMetLeuAsnValA  1330 1350 1370  AATGCTCTACCACAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGCATTGTGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGCATTGTGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGCATTGTGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGCATTGTGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGCATTGTGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGCATTGTGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGCATTGTGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGCATTGTGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACGTAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACGTAAGAAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACGTAAGAAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACGTAAGAAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACGTAAGAAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACAAAAATGGCCTCTCACCTGGGGCCATTGCTGGAACAAAAAAAA	STTGGA					
AATGCTCTACCACAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGCATTGTGA						
1390 1410 1430						
GTAGTGGCCCTGGTTGCTCTGATAGCAGTAGCCCTGGCATGTTTTCTGCATTTCC ValValAlaLeuValAlaLeuIleAlaValAlaLeuAlaCysPheLeuHisPheC						
1450 1470 1490						
ACCGGCAGCTCAGGACCACTCCAATGACCCACCTAACAAGATGAATGA	ACCGGCAGCTCAGGACCACTCCAATGACCCACCTAACAAGATGAATGA					
1510 1530 1550						
TACCCTGAACTTTGAAGCCCAGCAACCCACACCAACCTTCAGCCTCCCCATC	CTAAC					
1570 1590 1610						
AGCCACAGAAATAATTTATTCAGAAGTAAAAAAGCAGTAATGAAACCTGAAAAAA	LAAAAA					
1630						
35 AAAAAAAAA						
40						

55

45

50

^

# SEQUENCE AND TRANSLATION OF CDNA OF TM-4

5	10	30	50
٠	CAGCCGTGCTCGAAGCGTT	CCTGGAGCCCAAGCTCTCC	CCACAGGTGAAGACAGGGCCA
10	70	90	110
	GCAGGAGACACCATGGGGC MetGlyH	ACCTCTCAGCCCCACTTCAC IisLeuSerAlaProLeuHis	AGAGTSCGTGTACCCTGGCAG ArgValArgValProTrpGln
15	130 -	150	170
	GGGCTTCTGCTCACAGCCT GlyLeuLeuLeuThrAlaS	CACTTCTAACCTTCTGGAAC	CCGCCCACCACTGCCCAGCTC ProproThrThrAlaGlnLeu
20	190	210 <sup>-</sup>	230
	ACTACTGAATCCATGCCAT ThrThrGluSerMetProP	TCAATGTTGCAGAGGGGAAG heAsnValAlaGluGlyLys	GAGGTTCTTCTCCTTGTCCAC GluValLeuLeuLeuValHis
25	250	270	290
30	AATCTGCCCCAGCAACTTT AsnLeuProGlnGlnLeuP	TTGGCTACAGCTGGTACAAA heGlyTyrSerTrpTyrLys	GGGGAAAGAGTGGATGGCAAC GlyGluArgValAspGlyAsn
	310	330	350
35	CGTCAAATTGTAGGATATG ArgGlnIleValGlyTyrA	CAATAGGAACTCAACAAGCT laIleGlyThrGlnGlnAla	ACCCCAGGGCCCGCAAACAGC ThrProGlyProAlaAsnSer
	370	390	410
40	GGTCGAGAGACAATATACCGGlyArgGluThrIleTyrP	CCAATGCATCCCTGCTGATC roAsnAlaSerLeuLeuIle	CAGAACGTCACCCAGAATGAC GlnAsnValThrGlnAsnAsp
	430	450	470
45	ACAGGATTCTACACCCTACA ThrGlyPheTyrThrLeuG	AAGTCATAAAGTCAGATCTT lnValileLysSerAspLeu	GTGAATGAAGAAGCAACTGGA ValAsnGluGluAlaThrGly

ĺ

	490	~	510	530	
5	CAGTTCCATGTATACCC GlnPheHisValTyrPr	GGAGCTGC OGluLeuP	CCAAGCCC' COLYSPIO	TCCATCTCCAGCAACACTCCA SerileSerSerAsnAsnSerA	ACCCT SnPro
	550		570	590	•
10	GTGGAGGACAAGGATGC ValGluAspLysAspAl	TGTGGCCT' aValAlaPl	TCACCTGT( heThrCys)	GAACCTGAGACTCAGGACACAA GluProGluThrGlnAspThrT	CCTAC hrTyr
	610		630	650	
15	CTGTGGTGGATAAACAA LeuTrpTrpIleAsnAs	TCAGAGCC nGlnSecL	TCCCGGTC. euProVal	AGTCCCAGGCTGCAGCTGTCCA SerProArgLeuGlnLeuSerA	ATGGC snGly
	670		690	710	
20				AATGACACAGGACCCTATGAGT AsnAspThrGlyProTyrGluC	
	730		750	770	
25	ATACAGAACCCAGTGAG IleGlnAsnProValSe	TGCGAACC rAlaAsnA	GCAGTGAC rgSerAsp	CCAGTCACCTTGAATGTCACCT ProvalThrLeuAsnValThrT	ATGGC yrGly
30	790		810	830	
30				TATTACCGTCCAGGGGCAAACC TyrTyrArgProGlyAlaAsnL	
35	850		870	890	
				CAGTACTCCTGGCTTATCAATG GlnTyrSerTrpLeuIleAsnG	
40	910		930	950	
	TTCCAGCAAAGCACACA PheGlnGlnSerThrGl	AGAGCTCT nGluLeuP	TTATCCCT hellePco	AACATCACTGTGAATAATAGTG AsnIleThrValAsnAsnSerG	GATCC lyser
45	970 .		990	1010	
	TATACCTGCCACGCCAA TyrThrCysHisalaAs	NTAACTCAG InAsnSerV	TCACTGGC	TGCAACAGGACCACAGTCAAGA CysAsnArgThrThrValLysT	CGATC hrile
50					

1030 1050 1070 ATAGTCACTGATAATGCTCTACCACAAGAAAATGGCCTCTCACCTGGGGCCATTGCTGGC IleValThrAspAsnAlaLeuProGlnGluAsnGlyLeuSerProGlyAlaIleAlaGly 1090 1110 1130 ATTGTGATTGGAGTAGTGGCCCTGGTTGCTCTGATAGCAGTAGCCCTGGCATGTTTTCTG 10 IleValIleGlyValValAlaLeuValAlaLeuIleAlaValAlaLeuAlaCysPheLeu 1170 1150 1190 15 HisPheGlyLysThrGlySerSerGlyProLeuGln 12.10 1230 1250 20 1270 1290 1310 CCCATCCCTAAGAGCCACAGAAATAATTTATTCAGAAGTAAAAAAGCAGTAATGAAACCT 25 1330 GAAAAAAAAAAAAAAA 30

The present invention is also directed to a replicable recombinant cloning vehicle ("vector") having an insert comprising a nucleic acid, e.g., DNA, which comprises a base sequence which codes for a CEA peptide or a base sequence hybridizable therewith.

This invention also relates to a cell that is transformed/transfected, infected or injected with the above described replicable recombinant cloning vehicle or nucleic acid hybridizable with the aforementioned cDNA. Thus the invention also concerns the transfection of cells using free nucleic acid, without the use of a cloning vehicle.

Still further, the present invention concerns a polypeptide expressed by the above described transfected, infected or injected cell, which polypeptide exhibits immunological cross-reactivity with a CEA, as well as labelled forms of the polypeptide. The invention also relates to polypeptides having an amino acid sequence, i.e., synthetic peptides, or the expression product of a cell that is transfected, injected, infected with the above described replicable recombinant cloning vehicles, as well as labelled forms thereof. Stated otherwise, the present invention concerns a synthetic peptide having an amino acid sequence corresponding to the entire amino acid sequence or a portion thereof having no less than five amino acids of the aforesaid expression product.

The invention further relates to an antibody preparation specific for the above described polypeptide.

Another aspect of the invention concerns an immunoassay method for detecting CEA or a functional equivalent thereof in a test sample comprising

- (a) contacting the sample with the above described antibody preparation, and
- (b) determining binding thereof to CEA in the sample.

55

The invention also is directed to a nucleic acid hybridization method for detecting a CEA or a related nucleic acid (DNA or RNA) sample in a test sample comprising

- (a) contacting the test sample with a nucleic acid probe comprising a nucleic acid, which comprises a base sequence which codes for a CEA peptide sequence or a base sequence that is hybridizable therewith, and
- (b) determining the formation of the resultant hybridized probe.

The present invention also concerns a method for detecting the presence of carcinoembryonic antigen or a functional equivalent thereof in an animal or human patient in vivo comprising

- a) introducing into said patient a labeled (e.g., a radio-opaque material that can be detected by X-rays, radiolabeled or labeled with paramagnetic materials that can be detected by NMR) antibody preparation according to the present invention and
- b) detecting the presence of such antibody preparation in the patient by detecting the label.

In another aspect, the present invention relates to the use of an antibody preparation according to the present invention for therapeutic purposes, namely, attaching to an antibody preparation radionuclides, toxins or other biological effectors to form a complex and introducing an effective amount of such complex into an animal or human patient, e.g., by injection or orally. The antibody complex would attach to CEA in a patient and the radionuclide, toxin or other biological effector would serve to destroy the CEA expressing cell.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic representation of the transmembrane CEA's

#### **DETAILED DESCRIPTION OF THE INVENTION**

In the parent application 87111/68, published as EP-A-263 933, applicants described the following CEA's:

	ATCC No.
CEA-(a) partial CEA (pcLV7)	
CEA-(b) full coding CEA (pc 15LV7)	67709
CEA-(c) TM-1 (FL-CEA; pc 19-22)	67710
CEA-(d) NCA (pcBT 20)	67711

In the present application, applicants described the following CEA's:

	ATTC No.
CEA-(e) TM-2 (pc E22) CEA-(f) TM-3 (pc HT-6) CEA-(g) TM-4.	67712 67708

ATCC Nos. 67708, 67709, 67710, 67711 and 67712 were all deposited with the American Type Culture Collection on May 25, 1988.

30

5

15

20

25

35

50

The sequences for CEA-(a), CEA-(b), CEA-(c) and CEA-(d) are given hereinbelow:

CEA-(a):

(b)

10 20 30 40 50

C ACC ATG GAG TCT CCC TCG GCC CCT CTC CAC AGA TGG TGC ATC CCC TGG CAG AGG CTC Met Glu Ser Pro Ser Ala Pro Leu His Arg Trp Cys Ile Pro Trp Gln Arg Leu

CTG CTC ACA GCC TCA CTT CTA ACC TTC TGG AAC CCG CCC ACC ACT GCC AAG CTC ACT Leu Leu Thr Ala Ser Leu Leu Thr Phe Trp Asn Pro Pro Thr Thr Ala Lys Leu Thr ATT GAA TCC ACG CCG TTC AAT GTC GCA GAG GGG AAG GAG GTG CTT CTA CTT GTC CAC Ile Glu Ser Thr Pro Phe Asn Val Ala Glu Gly Lys Glu Val Leu Leu Val His 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 AAT CTG CCC CAG CAT CTT TIT GGC TAC AGC TGG TAC AAA GGT GAA AGA GTG GAT GGC Asn Leu Pro Gln His Leu Phe Gly Tyr Ser Trp Tyr Lys Gly Glu Arg Val Asp Gly 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 AAC CGT CAA ATT ATA GGA TAT GTA ATA GGA ACT CAA CAA GCT ACC CCA GGG CCC GCA Asn Arg Gln Ile Ile Gly Tyr Val Ile Gly Thr Gln Gln Ala Thr Pro Gly Pro Ala 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 TAC AGT GGT CGA GAG ATA ATA TAC CCC AAT GCA TCC CTG CTG ATC CAG AAC ATC ATC \_ Tyr Ser Gly Arg Glu Ile Ile Tyr Pro Asn Ala Ser Leu Leu Ile Gln Asn Ile Ile 61 62. 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 CAG AAT GAC ACA GGA TTC TAC ACC CTA CAC GTC ATA AAG TCA GAT CTT GTG AAT GAA Gln Asn Asp Thr Gly Phe Tyr Thr Leu His Val Ile Lys Ser Asp Leu Val Asn Glu 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 GAA GCA ACT GGC CAG TTC CGG GTA TAC CCG GAG CTG CCC AAG CCC TCC ATC TCC AGC Glu Ala Thr Gly Gln Phe Arg Val Tyr Pro Glu Leu Pro Lys Pro Ser Ile Ser Ser 99 101 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 AAC AAC TCC AAA CCC GTG GAG GAC AAG GAT GCT GTG GCC TTC ACC TGT GAA CCT GAG Asn Asn Ser Lys Pro Val Glu Asp Lys Asp Ala Val Ala Phe Thr Cys Glu Pro Glu 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136

	;	520		5	30		540				550			5	60			570	
	ACT C	AG GAC	GCA	¥CC	TAC	CTG	TGG	TGG	GTA	AAC	AAT	CAG .	AGC	стс	CCG	GTC	AGT	ССС	
5	Thr G	in Asp	Ala	Thr	Tyr	Leu	Тгр	Тгр	Val	Asn	Asn	Gln	Ser	Leu	Pro	Va1	Ser	Pro	
	137 1	38 139	140	141	14Z	143	144	145	146	147	148	149	150	יכו	132	123	134	133	
		580	)		S	90			600			610				620			
10	VCC C	TG CAG	CTG	TCC	AAT	- -	<b>ΔΔ</b> Γ	AGG	ACC	стс	ACT	e CTA	TTC	AAT	GTC	ACA	AGA	AAT	
	Arg L	eu Gln	Leu	Ser	Asn	Gly	Asn	Arg	Thr	Leu	Thr	Leu	Phe	Asn	41	Thr	Arg	Asn	l
	156 1	57 158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	,
15	630		640	)		(	50			660			670	)			680		
	GAA C	AA GCA	AGC	TAC	AAA	TGT	GAA	ACC	CAG	AAC	CCA	GTG	AGT	GCC	AGG	CGC	AGT	GAT	
		In Ala																	
••	175 1	76 177	1/8	1/9	180	181	182	183	154	183	100	16/	100	103	150	171	1 34	. 193	•
20					_			·							_			740	
	6	90		70	3			710			720			73 •				740	
		TC ATC																	
25		/al Ile 195 196																	
20	194	ספו כפו	197	190	133	200	201	202	203	204	203	200	20,	200					•
		750			76	0			770			780			79	0			
	464	•		754	•		447		•	CTC	TCC	•		· cc	•	•	T AA	ר רר	Δ
30		• CT TAC	AGA		GGG	GAA		CTG	- AAC			- TGC	CAC		GC(	C TC			
30	Thr S	•	: AGA : Arg	Ser	GGG G1y	GAA G1u	Asn	CTG Leu	# AAC Asn	Leu	Ser	TGC Cys	CAC His	A1a	GC0	TC' Se	r As	n Pr	0
	Thr S	CT TAC Ser Tyr	: AGA : Arg	Ser 217	GGG G1y	GAA G1u	Asn 220	CTG Leu	AAC Asn 222	Leu	Ser	TGC Cys	CAC His 226	A12	GC0	TC Se ZZ	r As	n Pr	0
30 35	7hr 5 213 2 800	ETT TAC Ser Tyr 214 215	AGA Arg 216	Ser 217	GGG G1y 218	GAA G1u 219	Asn 220 0	CTG Leu 221	AAC Asn 222	Leu 223 830	Ser 224	TGC Cys 225	CAC His 226	A14 227	GC( A1: 228	7 TC 1 Se 2 ZZ	r As 9 23 50	n Pr 0 23	0
	800	TCT TAC Ser Tyr 214 215 GCA CAG	AGA Arg 216 810 **	Ser 217 TCT	GGG G1y 218	GAA GIu 219 82 87	Asn 220 0 GTC Val	CTG Leu 221	AAC Asn 222	Leu 223 830 ACT	Ser 224	TGC Cys 225	840 840 840	A1a 227	A GCC A A 1 a 2 2 2 8	Ser	F As 9 23 50 = A GA n G1	n Pro 0 23 G CT w Le	0 1
	800	ECT TAC Ser Tyr 214 215	AGA Arg 216 810 **	Ser 217 TCT	GGG G1y 218	GAA GIu 219 82 87	Asn 220 0 GTC Val	CTG Leu 221	AAC Asn 222	Leu 223 830 ACT	Ser 224	TGC Cys 225	840 840 840	A1a 227	A GCC A A 1 a 2 2 2 8	Ser	F As 9 23 50 = A GA n G1	n Pro 0 23 G CT w Le	0 1
	800	FIGURE 14 CONTRACT TACK	AGA Arg 216 810 **	Ser 217 TCT	GGG G1y 218	GAA GIu 219 82 87	Asn 220 0 GTC Val	CTG Leu 221	AAC Asn 222	Leu 223 830 ACT	Ser 224	TGC Cys 225	840 840 840	TC0	A GCC A A1a 7 228	Ser	F As 9 23 50 = A GA n G1 8 24	G CT u Le 9 25	0 1
35	800	TCT TAC Ser Tyr 214 215 GCA CAG	AGA Arg 216 810 **	Ser 217 TCT	GGG G1y 218 TGG Trp	GAA GIu 219 82 87	Asn 220 0 GTC Val	CTG Leu 221 AAT Asn 240	AAC Asn 222	Leu 223 830 ACT	Ser 224	TGC Cys 225	840 840 840	A1a 227	A GCC A A1a 7 228	Ser	F As 9 23 50 = A GA n G1 8 24	n Pro 0 23 G CT w Le	0 1
35	Thr 5 213 2 800	GCA CAGATA GTM 233 234	810 = 1 TAC	Ser 217	GGG Gly 218	82 82 82 6 TTT Phe 238	Asn 220 0 GTC Val 239 88	CTG Leu ZZ1 AAT Asn 240	AAC Asn ZZZZ	830 - ACT Thr 242	TTC: Phe: 243	TGC Cys 225 CAG Gln 244	CAC His 2266 840	900 ±	ACC ACC Thi	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	7 As 23 550 = A GA G	G CT Le 9 25	0 1 C U U
35	800 ** CCT ( Pro / 232 :	GCA CAGATA GIN 233 234	810 = TAC	Ser   217	GGGG Gly 218 TGG Trp 237	82 82 3 TTT 9 Phe 238	Asn 2200 GTC Val 2399 88	CTG Leu 221 AAT Asn 240	AACC Asn ZZZZ	223 830 ACT Thr 242	Ser   224	TGC Cys ZZ5 CAG G1n 244	CAC His 226 846 # CAA GIT ACC Thi	900 F TGG	ACC: ACC: ACC: This 5 24'	8 22: 8 C CA 7 24	7 As 23 550 = A GA G	G CT Leg 25	o 1 C u 0 C
35	800 ** CCT ( Pro / 232 :	GCA CAGATA GTM 233 234	810 = TAC	Ser   217	GGGG Gly 218 TGG Trp 237	82 82 3 TTT 9 Phe 238	Asn 2200 GTC Val 2399 88	CTG Leu 221 AAT Asn 240	AACC Asn ZZZZ	223 830 ACT Thr 242	Ser   224	TGC Cys ZZ5 CAG G1n 244	CAC His 226 846 # CAA GIT ACC Thi	900 F TGG	ACC: ACC: ACC: This 5 24'	8 22: 8 C CA 7 24	7 As 23 550 = A GA G	G CT Leg 25	o 1 C u 0 C
35 40	800 ** CCT ( Pro / 232 :	GCA CAGATA GIN 233 234	810 = TAC	Ser   217	GGGG Gly 218 TGG Trp 237	82 82 83 6 TTT 9 Phe 238 7 GTO 7 Val	Asn 2200 GTC Val 2399 88	CTG Leu 221 AAT Asn 240	GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	223 830 ACT Thr 242	Ser   224	TGC Cys ZZ5 CAG G1n 244	CAC His 226 846 # CAA GIT ACC Thi	900 F TGG	ACC: ACC: ACC: This 5 24'	8 See 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	7 As 23 550 = A GA G	G CT Le 25	0 1 C u 0 0
35 40	Thr 9 213 2 800 # CCT ( Pro 232 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	GCA CAG ATC CCC TILE Pro 233 234  ATC CCC TILE Pro 252 253	810 = 1 TAC	Ser   217	GGG G1y 218 TGG Trp 237	82 82 83 85 77 6 77 6 77 6 77 6 77 6 77 6 77 6 7	Asn 2200 0 GTC Val 3 239 88 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	CTG Leu 2211  AAT Asn 240  AAT Ass 255	AACC Asn ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ	223 830 ACT Thr 242 GGA GGA	Ser   224	TGC Cys 225 CAG Gin 244	840 ************************************	900 = 900 = 0 A GT	GC ACC This 5 24'	8 Se	F AS 23 550 F A GA G	G CT Leg 25	C C C C C C C C C C C C C C C C C C C
35 40	Thr S 213 2 800 # CCT ( Pro 232 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	GCA CAGATA GIN 233 234  ATC CCC 11e Pro 252 253	810 810 6 TAC 1 Tyr 1 235 2 AAC 1 GGC 1 GGC	Ser   217	GGG G1y 218 TGG Trp 237	SGAA GTU 219 82 87 6 TTT 9 Phe 7 238	Asn 2200 GTC Val 3 239 88 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	CTG Leu 221 AAT Asn 240 AS AS AS AS AS AS AS AS AS AS AS AS AS	AACC Asn ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ	E ACC   The Control of the Control o	Ser   224	TGC Cys 225 CAG G1n 244 TAT Tyr 263 950 - G ATC TITE THE THE THE THE THE THE THE THE THE T	EACH His 226 840 a GTr 245 ACC ACC Th	900 = 900 = 0	GCC ACC This 5 24"	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	F AS 23 550 = A GA G	G CT Le Pi AG CC Lu Pi AG CC L	C C C C C C C C C C C C C C C C C C C

CCC AAA CCC TTC ATC ACC AGC AAC AAC TCC AAC CCC GTG GAG GAT GAG GAT GCT GTA Pro Lys Pro Phe I1e Thr Ser Asn Asn Ser Asn Pro Val Glu Asp Glu Asp Ala Val 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307  1030 1040 1050 1060 1070 1080  1030 1040 1050 1060 1070 1080  1030 1040 1050 1060 1070 1080  1030 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 320 321 322 323 324 325 326 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 320 321 322 323 324 325 326 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 320 321 322 323 324 325 326 320 320 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 320 321 322 323 324 325 326 320 320 320 320 320 320 320 320 320 320	5	Pro			TTC			•			•				•					
289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307  1030 1040 1050 1060 1070 1080  GCC TTA ACC TGT GAA CCT GAG ATT CAG AAC ACA ACC TAC CTG TGG TGG GTA AAT AAT A1A Leu Thr Cys Glu Pro Glu Ile Gln Asn Thr Thr Tyr Leu Trp Trp Val Asn Asn 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326  15 1090 1100 1110 1120 1130 1140  CAG AGC CTC CGG GTC AGT CCC AGG CTG CAG CTG TCC AAT GAC AAC AGG ACC CTC ACT GIN Ser Leu Pro Val Ser Pro Arg Leu Gln Leu Ser Asn Asp Asn Arg Thr Leu Thr 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345  20 1150 1160 1170 1180 1190  CTA CTC AGT GTC ACA AGG AAT GAT GTA GGA CCC TAT GAG TGT GGA ATC CAG AAC CAA Leu Leu Ser Val Thr Arg Asn Asp Val Gly Pro Tyr Glu Cys Gly Ile Gln Asn Glu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364  1200 1210 1220 1230 1240 1250  30 TTA AGT GTT GAC CAC AGG GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC CAC Leu Ser Val Asp His Ser Asp Pro Val Ile Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  35 1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40 1320 1330 1340 1350 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Por Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Por Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Por Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Por Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Por Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Por Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Por Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His	5		Lys			ATC	ACC	AGC	AAC	AAC	TCC	AAC	CCC	GTG	GAG	GAT	GAG	GAT	GCT	GTA
1030 1040 1050 1060 1070 1080  GCC TTA ACC TGT GAA CCT GAG ATT CAG AAC ACA ACC TAC CTG TGG TGG GTA AAT AAT AT AT AT Leu Thr Cys GTU Pro GTU TIE GTN ASN THE Thr Tyr Leu Trp Trp Val ASN ASN 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326  15 1090 1100 1110 1120 1130 1140  CAG AGC CTC CGG GTC AGT CCC AGG CTG CAG CTG TCC AAT GAC AAC AGG ACC CTC ACT GTN Ser Leu Pro Val Ser Pro Arg Leu GTN Leu Ser Asn Asp Asn Arg Thr Leu Thr 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345  20  TTA CTC AGT GTC ACA AGG AAT GAT GTA GGA CCC TAT GAG TGT GGA ATC CAG AAC GAA Leu Leu Ser Val Thr Arg Asn Asp Val GTy Pro Tyr GTu Cys GTy Tle GTN Asn GTu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364  1200 1210 1220 1230 1240 1250  TTA AGT GTT GAC CAC AGC GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC GAC Leu Ser Val Asp His Ser Asp Pro Val Tte Leu Asn Val Leu Tyr GTy Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  35 1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr T1e Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro GTy Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40 1320 1330 1340 1350 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Atla Ala Ser Asn Pro Pro Ala GIn Tyr Ser Trp Leu IIe Asp GTy Asp IIe Cys His Atla Ala Ser Asn Pro Pro Pro Ala GIn Tyr Ser Trp Leu IIe Asp GTy Asn IIe CYs His Atla Ala Ser Asn Pro Pro Pro Ala GIn Tyr Ser Trp Leu IIe Asp GTy Asp IIe CYs His Atla Ala Ser Asn Pro Pro Ala GIn Tyr Ser Trp Leu IIe Asp GTy Asn IIe CAD Add Add Add Add Add Add Add Add Add A		259	200																	
GCC TTA ACC TGT GAA CCT GAG ATT CAG AAC ACA ACC TAC CTG TGG TGG GTA AAT AAT AAT AAT AAT Leu Thr Cys Glu Pro Glu I1e Gln Asn Thr Thr Tyr Leu Trp Trp Val Asn Asn 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326  15 1090 1100 1110 1120 1130 1140  CAG AGC CTC CCG GTC AGT CCC AGG CTG CAG CTG TCC AAT GAC AAC AGG ACC CTC ACT GIn Ser Leu Pro Val Ser Pro Arg Leu Gin Leu Ser Asn Asp Asn Arg Thr Leu Thr 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345  20 1150 1160 1170 1180 1190  CTA CTC AGT GTC ACA AGG AAT GAT GTA GGA CCC TAT GAG TGT GGA ATC CAG AAC AGA AC Leu Leu Ser Val Thr Arg Asn Asp Val Gly Pro Tyr GTu Cys Gly Tle Gln Asn GTu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364  1200 1210 1220 1230 1240 1250  1200 1210 1220 1230 1240 1250  30 TTA AGT GTT GAC CAC AGG GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC GAC Leu Ser Val Asp His Ser Asp Pro Val Tle Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  35 1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Tle Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40 1320 1330 1340 1350 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu IIe Asp Gly Asn IIe			290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307
Ala Leu Thr Cys Glu Pro Glu IIe Gln Asn Thr Thr Tyr Leu Trp Trp Val Asn Asn 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 328 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 328 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 328 329 330 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 326 326 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 320 320 321 322 323 324 325 326 333 334 335 336 337 338 339 340 341 342 343 344 345 320 320 320 321 322 323 324 325 326 333 334 335 336 337 338 339 340 341 342 343 344 345 320 320 320 321 322 323 324 325 326 327 328 329 340 341 342 343 344 345 345 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 326 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 326 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 326 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 326 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 385 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 340 340 340 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 340 340 465 466 466 467 468 468 468 468 468 468 468 468 468 468			)		10	_							)		10	70		1		
308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326  1090 1100 1110 1120 1130 1140  CAG AGC CTC CCG GTC AGT CCC AGG CTG CAG CTG TCC AAT GAC AAC AGG ACC CTC ACT GIn Ser Leu Pro Val Ser Pro Arg Leu Gin Leu Ser Asn Asp Asn Arg Thr Leu Thr 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345  20  1150 1160 1170 1180 1190  CTA CTC AGT GTC ACA AGG AAT GAT GAG GCC TAT GAG TGT GGA ATC CAG AAC GAA Leu Leu Ser Val Thr Arg Asn Asp Val Gly Pro Tyr Glu Cys Gly Tle Gin Asn Glu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364  1200 1210 1220 1230 1240 1250  30 TTA AGT GTT GAC CAC AGC GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC GAC Leu Ser Val Asp His Ser Asp Pro Val Tle Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  35 1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Tle Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Tle Asp Gly Asn Tel Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Tle Asp Gly Asn Tel Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Tle Asp Gly Asn Tel Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Tle Asp Gly Asn Tel Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Tle Asp Gly Asn Tel Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Tle Asp Gly Asn Tel Cys His Asn Can Cha	10																			
15 1090 1100 1110 1120 1130 1140  CAG AGC CTC CCG GTC AGT CCC AGG CTG CAG CTG TCC AAT GAC AAC AGG ACC CTC ACT GIn Ser Leu Pro Val Ser Pro Arg Leu Gin Leu Ser Asn Asp Asn Arg Thr Leu Thr 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345  20  1150 1160 1170 1180 1190  CTA CTC AGT GTC ACA AGG AAT GAT GTA GGA CCC TAT GAG TGT GGA ATC CAG AAC GAA Leu Leu Ser Val Thr Arg Asn Asp Val Gly Pro Tyr Glu Cys Gly Ile Gin Asn Glu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364  1200 1210 1220 1230 1240 1250  30 17A AGT GTT GAC CAC AGC GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC GAC Leu Ser Val Asp His Ser Asp Pro Val Tle Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  35 1260 1270 1280 1290 1300 1310  CCC ACC ACT ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Pro Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile Cys His Ala Ala Ser Asn Pro Pro Ala Gin Tyr Ser Trp Leu Ile Asp Gly Asn Ile					-									•		•				
CAG AGC CTC CCG GTC AGT CCC AGG CTG CAG CTG TCC AAT GAC AAC AGG ACC CTC ACT GIn Ser Leu Pro Val Ser Pro Arg Leu Gin Leu Ser Asn Asp Asn Arg Thr Leu Thr 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 20    1150					•••		•••	•			• • •		•.,	544				-	-	525
Gin Ser Leu Pro Val Ser Pro Arg Leu Gin Leu Ser Asn Asp Asn Arg Thr Leu Thr 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 20  1150 1160 1170 1180 1190  CTA CTC AGT GTC ACA AGG AAT GAT GTA GGA CCC TAT GAG TGT GGA ATC CAG AAC GAA Leu Leu Ser Val Thr Arg Asn Asp Val Gly Pro Tyr Glu Cys Gly Ile Gin Asn Glu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 363 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 364 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 365 365 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 365 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 365 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 365 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 402 402 402 402 402 402 402 402 402	15		1090	)		11	00		1	1110			1120	)		11	30		•	
20  1150 1160 1170 1180 1190  CTA CTC AGT GTC ACA AGG AAT GAT GTA GGA CCC TAT GAG TGT GGA ATC CAG AAC GAA Leu Leu Ser Val Thr Arg Asn Asp Val Gly Pro Tyr Glu Cys Gly Ile Gln Asn Glu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364  1200 1210 1220 1230 1240 1250  TTA AGT GTT GAC CAC AGC GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC GAC Leu Ser Val Asp His Ser Asp Pro Val Ile Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  25  1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40  1320 1330 1340 1350 1360 1360 1360 137 138 379 384 385 GGG AAC ATC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC CYs His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile COS AGC																				
20  1150 1160 1170 1180 1190  CTA CTC AGT GTC ACA AGG AAT GAT GTA GGA CCC TAT GAG TGT GGA ATC CAG AAC GAA Leu Leu Ser Val Thr Arg Asn Asp Val Gly Pro Tyr Glu Cys Gly Ile Gln Asn Glu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364  1200 1210 1220 1230 1240 1250  30  TTA AGT GTT GAC CAC AGC GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC GAC Leu Ser Val Asp His Ser Asp Pro Val Ile Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  35  1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40  1320 1330 1340 1350 1360 136 CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile 403 404 405 405 405 405 405 405 405 405 405																				
CTA CTC AGT GTC ACA AGG AAT GAT GTA GGA CCC TAT GAG TGT GGA ATC CAG AAC GAA Leu Leu Ser Val Thr Arg Asn Asp Val Gly Pro Tyr Glu Cys Gly Ile Gln Asn Glu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 362 362 363 364 362 363 364 362 363 364 362 363 364 362 363 364 362 363 364 362 363 364 362 363 364 362 363 364 362 363 364 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 383 365 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 365 365 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 365 365 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 385 385 385 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 400 400 400 400 400 400 400 400 400		11																		
Leu Leu Ser Val Thr Arg Asn Asp Val Gly Pro Tyr Glu Cys Gly Ile Gln Asn Glu 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364  1200 1210 1220 1230 1240 1250  TTA AGT GTT GAC CAC AGC GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC GAC Leu Ser Val Asp His Ser Asp Pro Val Ile Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile 403 404 405 405 407 409 400 411 412 417 417 416 456 477 478 410 420 411				1150	0		1								)		11	90		
25  346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364  1200 1210 1220 1230 1240 1250  37  38  39  30  30  30  30  30  30  30  30  30																				
1200 1210 1220 1230 1240 1250  TTA AGT GTT GAC CAC AGC GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC GAC Leu Ser Val Asp His Ser Asp Pro Val Ile Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile	25																			
TTA AGT GTT GAC CAC AGC GAC CCA GTC ATC CTG AAT GTC CTC TAT GGC CCA GAC GAC Leu Ser Val Asp His Ser Asp Pro Val Ile Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile						•														
Leu Ser Val Asp His Ser Asp Pro Val Ile Leu Asn Val Leu Tyr Gly Pro Asp Asp 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile		1200			1210	)		12	220		•	1230			1240	)		12	250	
365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383  1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr I1e Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu I1e Asp Gly Asn I1e	30																			
1260 1270 1280 1290 1300 1310  CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile		365	366	367	368	369	370	371	372	373	11e	Leu 375	Asn 376	Val 377	1eu	Tyr 379	G1y 380	Pro 381	Asp 382	Asp 383
CCC ACC ATT TCC CCC TCA TAC ACC TAT TAC CGT CCA GGG GTG AAC CTC AGC CTC TCC Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile																				
Pro Thr Ile Ser Pro Ser Tyr Thr Tyr Tyr Arg Pro Gly Val Asn Leu Ser Leu Ser 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile	35	1				1270	)		12	280		1	290			1300	)		1:	310
384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402  40  1320  1330  1340  1350  1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile																				
1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC  Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile																				
1320 1330 1340 1350 1360  TGC CAT GCA GCC TCT AAC CCA CCT GCA CAG TAT TCT TGG CTG ATT GAT GGG AAC ATC  Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile	40					-											-			
Cys His Ala Ala Ser Asn Pro Pro Ala Gln Tyr Ser Trp Leu Ile Asp Gly Asn Ile	40		1					)		13			` 1					)		
403 404 405 406 407 409 400 410 412 412 414 415 415 417 419 410 420 421																				
40	45																			
	45													•••	•	•••	•••			
1370 1380 1390 1400 1410 1420		1370		1	380				)		14			1	410				)	
CAG CAA CAC ACA CAA GAG CTC TTT ATC TCC AAC ATC ACT GAG AAG AAC AGC GGA CTC	50																			
Gln Gln His Thr Gln Glu Leu Phe Ile Ser Asn Ile Thr Glu Lys Asn Ser Gly Leu 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440																				

	1430	)	1	440			145	0		1	460			1470			1480	)
				•			•				•			•			•	
	TAT AC																	
5	Tyr Th 441 44																	
				<b>3</b> .			0	,	730	731	432	733	7.54	433	430	731	730	439
	ı	1490		1	500			1510	)		15	520		,	1530			1540
		*			•			•				•			•			•
10	ATC AC																	
	Ile Th 460 46																	
			-03		70,	400	707	700	707	770	7,,	7/2	7/3	7/7	7/3	7/0	7//	7/0
15		15	550		1	1560			1570	)		19	580		1	590		
15			•			•			•				•			•		
	GTG GA	G GAC	AAG	GAT	GCT	GTG	$\infty$	TTC	ACC	TGT	GAA	CCT	GAG	GCT	CAG	AAC	ACA	ACC
	Val G1	0 ASP	Lys	ASP	Ala	Val	Ala	Phe	Thr	Cys	Glu	Pro	Glu	Ala	Gla	Asn	Thr	Thr
	479 48	.0 461	402	403	404	400	400	487	488	489	490	491	492	493	494	495	496	497
20																		
	1600		16	10			1620			163	)		10	540			1650	
	TAC CT	G TGG	TGG	GTA	ΔΔΤ	GET	CAG	AGC	CTC		GTC	ACT	ccc	ACC	CTC	CAC		TCC
	Tyr Le	u Trp	Trp	Val	Asn	Glv	Gln	Ser	Leu	Pro	Val	Ser	Pro	Ara	Leu	Gla	Leu	Ser
	498 49	9 500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516
25																		
	16	60		16	70		,	1680			169	<b>1</b>		1.	700			1710
		•			•			*				•		•	-			.,
	AAT GO																	
30	Asn G1																	
	517 51	8 519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535
		1720				• • • •							_					
		1720	,		17	730			1740			1750	כ		17	760		
35	GTA TO	T GGA	ATC	CAG	AAC	TCA	GTG	AGT	GCA	AAC	CGC	AGT	GAC	CCA	GTC	ACC	CTG	GAT
	Val Cy																	
	536 53	7 538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554
	1770		1700			_								_		_		
40 .	1770		1780			1,	790`			1800			1810	) .		18	320	
	GTC CT	C TAT	GGG	CCG	GAC	ACC	ccc	ATC	ATT	TCC	ככר	CCA	GAC	TCG	TCT	TAC	CTT	TCG
	Val Le																	
	555 55	6 557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573
45														•				
	183	0		1840			18	350			860			1870	)		18	880
	•			•				•			•			•				-
	GGA GC																	
	Gly Al																	
50	574 57	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592

			1890		: ,,	190	0		1	910			1920			1930	)		
	TGG	CGT	ATC	AAT	GGG	ATA	CCG	CAG	CAA	CAC	ACA	CAA	GTT	СТС	TTT	ATC	GCC	AAA	ATC
_	Тгр	Arg	Ile	Asn	Gly	He	Pro	Gln	Gln	His	Thr	Gln	Val	Leu	Phe	Ile	Ala	Lys	Ile
5	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610 	611
	1940			1950			196	0		11	970			1980			1990	n	
	•			•			•	-	٠	•	•			•			•	•	
10												GTC							
																			Arg
	612	613	514	015	510	617	618	619	620	621	622	623	624	625	626	627	628	629	630
	2	000		;	2010			202	0		20	, 030		;	2040			205	0
15	•,•	•			•			•							-			•	
												GCA							
																			Leu
	031	032	623	634	635	636	637	638	639	640	641	642	643	644	645	645	647	648	649
20										_									
		2	060			2070			208	0		20	90		;	2100			2110
	TCA	GCT	GGG	GCC	ACT	GTC	GGC	ATC	ATG	ATT	GGA	GTG	CTG	GTT	GGG	GTT	GCT	CTG	ATA
	Ser	Ala	Gly	Ala	Thr	Val	Gly	He	Het	Ile	Gly	Val	Leu	Val	Gly	Val	Ala	Leu	Ile
25	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668
			2	120		;	2130			2140	)		2	150		;	2160		
	TAG	CAG	ccc	TGG	TGT	AGT	TTC	TTC	ATT	TCA	GGA	AGA	CTG	ACA	GTT	GTT	TTG	CTT	СТТ
30																			
	217	0		2	180		;	2190			220	)		2	210		;	2220	
	CCT	TAA	AGC	ATT	TGC	AAC	AGC	TAC	AGT	CTA	AAA	TTG	CTT	СТТ	TAC	CAA	GGA	TAT	TTA
35																			
		223	'n		2.	240			2250			226	•		9	270			2200
		223	,		2.	•		•	-			2260	,		2.	270			2280
	CAG	AAA	ATA	CTC	TGA	CCA	GAG	ATC	GAG	ACC	ATC	CTA	GCC	AAC	ATC	GTG	AAA	ccc	CAT
40								-	•				-						
			2290	3		23	300		;	2310			2320	0		2:	330		
	CTC	TAC	TAA	AAA	TAC		ÄAT	GAG	CTG	GGC	TTG	GTG	GCG	CGC	ACC	TGT	AGT	ССС	AGT
45	2340			2350	<b>)</b> -		23	360		1	2370			2380	0		2:	3 <b>9</b> 0	
	•		•	•				•			•			•				•	
	TAC	TCG	GGA	GGC	TGA	GGC	AGG	AGA	ATC	GCT	TGA	ACC	CGG	GAG	GTG	GAG	ATT	GCA	GTG

	7	2400			2410	)		24	120		2	2430			2440			24	150
	AGC	CCA	GAT	CGC	ACC	ACT	GCA	стс	CAG	TCT	GGC	AAC	AGA	GCA	AGA	стс	CAT		
5		2	2460			2470				480		;	2490			2500			
	AAG	AAA	AGA	AAA	GAA	GAC	TCT	GAC	CTG	TAC	тст	TGA			GTT	TCT	GAT	ACC	ACT
10	2510		į	2520			253	0		25	540		;	2550			2560		
	GCA	CTG	TCT	GAG	AAT	TTC	CAA	AAC	TTT	AAT	GAA	CTA	ACT	GAC	AGC	TTC	ATG	AAA	CTG
15	25	570 *		:	2580			2590	)		20	500		;	2610			2620	
	TCC	ACC	AAG	ATC	AAG	CAG	AGA	AAA	TAA	TTA	ATT	TCA	TGG	GGA	CTA	AAT	GAA	CTA	ATG
20		26	530		;	2640 •			2650			20	6 <b>6</b> 0		 - (	2670			2680
	AGG	ATA	ATA	TTT	TCA	TAA	TTT	TTT	ATT	TGA	AAT	TTT	GCT	GAT	TCT	TTA	AAT	GTC	TTG
25			2	690 <sup>-</sup>		;	2700			2710			2	720		i	2730		
	TTT	ccc	AGA	τττ	CAG	GAA	ACT	TTT	ΤΤΤ			AGC	TAT		СТС	TTA		CAA	TTT
30	2740	)			750		;	2760			277	0		2	780		;	2790 <del>-</del>	
	GAT	AAA	ATA	TAC	TTT	TGT	GAA	CAA	AAA	TTG	AGA	CAT	TTA	CAT	TTT	ATC	ССТ	ATG	TGG
		2800	)		21	810		:	2820			283	0						
35	TCG	стс	CAG	ACT	TGG	GAA	ACT	ATT	CAT	GAA	TAT	TTA	TAT	TGT	ATG				
												٠							
40																			
<b>4</b> 5																			

	CEA-(c):		
5			
	10	30	50
10	CAGCCGTGCTCGAAGCGT	TCCTGGAGCCCAAGCTCTCCT	CCACAGGTGAAGACAGGGCCA
	70	90	110
15			AGAGTGCGTGTACCCTGGCAG ArgValArgValProTrpGln
	130	150	170
20			CCGCCCACCACTGCCCAGCTC
	190	210	230
25			GGAGGTTCTTCTCCTTGTCCAC
	250	270	290
30			AGGGGAAAGAGTGGATGGCAAC GlyGluArgValAspGlyAsn
35	310	330	350
33			ACCCCAGGGCCCGCAAACAGC ThrProGlyProAlaAsnSer
40	370	390	410
			CAGAACGTCACCCAGAATGAC GlnAsnValThrGlnAsnAsp
45	430	450	470
			GTGAATGAAGAAGCAACTGGA iValasnGluGluAlaThrGly

(2)

5	10	30	50
	CAGCCGTGCTCGAAGCGTT	CCTGGAGCCCAAGCTCTCCT	CCACAGGTGAAGACAGGGCCA
10	70	90	110
	GCAGGAGACACCATGGGGG MetGlyH	CACCTCTCAGCCCCACTTCAC	CAGAGTGCGTGTACCCTGGCAG
15	130	150	170
00	GGGCTTCTGCTCACAGCCTGlyLeuLeuLeuThralas	CACTTCTAACCTTCTGGAA( SerLeuLeuThrPheTrpAsi	CCCGCCCACCACTGCCCAGCTC PProProThrThrAlaGlnLeu
20	190	210	230
25	ACTACTGAATCCATGCCAT ThrThrGluserMetPro	TTCAATGTTGCAGAGGGGAAG	GGAGGTTCTTCTCCTTGTCCAC sGluValLeuLeuLeuValHis
	250	270	290
30	AATCTGCCCCAGCAACTTT AsnLeuProGloGloLeus	TTTGGCTACAGCTGGTACAA PheGlyTyrSerTrpTyrLy:	AGGGGAAAGAGTGGATGGCAAC sGlyGluArgValAspGlyAsn
	310	330	350
35	CGTCAAATTGTAGGATATG ArgGlnIleValGlyTyr	CAATAGGAACTCAACAAGCAAGCAAGCAAGCAAGCAAGCA	TACCCCAGGGCCCGCAAACAGC ThrProGlyProAlaAsnSer
	370	390	410
40	GGTCGAGAGACAATATACO GlyArgGluThrIleTyrI	CCCAATGCATCCCTGCTGAT ProAsnalaSerLeuLeuIl	CCAGAACGTCACCCAGAATGAC eGlnAsnValThrGlnAsnAsp
45			

55

	430	450	470	
5	ACAGGATTCTACACCCTAC ThrGlyPheTyrThrLeuG	AAGTCATAAAGTCAGATCTT lnValileLysSerAspLeu	GTGAATGAAGAAGCAACTGG ValAsnGluGluAlaThrGl	SA Ly
	490	510	530	
10	CAGTTCCATGTATACCCGG GlnPheHisValTyrProC	AGCTGCCCAAGCCCTCCATC	TCCAGCAACAACTCCAACC SerSerAsnAsnSerAsnP	CT ro
	550	570	590	
15	GTGGAGGACAAGGATGCTC ValGluAspLysAspAlaN	STGGCCTTCACCTGTGAACCT	GAGACTCAGGACACAACCT	AC yr
	610	630	650	
20	CTGTGGTGGATAAACAAT( LeuTrpTrpIleAsnAsn(	CAGAGCCTCCCGGTCAGTCCC	CAGGCTGCAGCTGTCCAATG DArgLeuGlnLeuSerAsnG	J.À
	670	690	710	
25	AACAGGACCCTCACTCTA AsnArgThrleuThrleu	CTCAGTGTCACAAGGAATGA( LeuSerValThrArgAsnAs)	CACAGGACCCTATGAGTGTG  oThrGlyProTyrGluCysG	AA Slu
30	730	750	770	
	ATACAGAACCCAGTGAGT IleGlnAsnProValSer	GCGAACCGCAGTGACCCAGT AlaAsnArgSerAsp?roVa	CACCTTGAATGTCACCTATO	31 À 20 C
35 ·	790	810	830	
	CCGGACACCCCCACCATT ProAspThr?roThrIle	TCCCCTTCAGACACCTATTA SerProSerAspThrTyrTy	CCGTCCAGGGGCAAACCTC rArgProGlyAlaAsnLeu!	AGC Ser
40		·		
45				

	550	870	890
5			CTCCTGGCTTATCAATGGAACA rSerTrpLewIleAsnGlyThr
3	910	930	950
10			CACTGTGAATAATAGTGGATCC eThrValAsnAsnSerGlySer
	970	990	1010
15	TATACCTGCCACGCCAAT TyrThrCysHisAlaAsn	AACTCAGTCACTGGCTGCAA AsnSerValThrGlyCysAs	CAGGACCACAGTCAAGACGATC
	1030	1050-	1070
20			AATCAAAGCCAGCAAGACCACA nileLysAlaSerLysThrThr
	1090	1110	1130
25			CACAAATGACACTGGAATCTCC rThrAsnAspThrGlyIleSer
30	1150	1170	1190
	ATCCGTTGGTTCTTCAAA IleArgTrpPhePheLys	AACCAGAGTCTCCCGTCCTC AsnGlnSerLeuProSerSe	GGAGAGGATGAAGCTGTCCCAG rGluArgMetLysLeuSerGln
35	1210	1230	1250
	GGCAACACCACCCTCAGC GlyAsnTh:Th:LeuSer	TATAAACCCTGTCAAGAGGGA TleAsnProValLysArgGl	GGATGCTGGGACGTATTGGTGT UASPAlaGlyThrTyrTrpCys
40			
45			
50			

	1270	1290	1310	
5	GAGGTCTTCAACCCAATCA GluValPheAsnProile	AGTAAGAACCAAAGCGACCG SerLysasnGlnSerAspPi	CCATCATGCTGAACGTAAACTA rOIleMetLeuAsnValAsnTy	r r
	1330	1350	1370	
10	AATGCTCTACCACAAGAA AsnalaLeuFroGlnGlu	AATGGCCTCTCACCTGGGGGAAAAAGGCTGGGGGGAAAAAAAA	CCATTGCTGGCATTGTGATTGG lailealaGlyileValileGl	A Y
	1390	1410	1430	
15	GTAGTGGCCCTGGTTGCT ValValAlaLeuValAla	CTGATAGCAGTAGCCCTGG LeuIleAlaValAlaLeuA	CATGTTTTCTGCATTTCGGGAA laCysPheLeuHisPheGlyLy	G S
	1450	1470	1490	
20	ACCGGCAGCTCAGGACCAG ThrGlySerSerGlyPro	CTCCAATGACCCACCTAAC LeuGln	AAGATGAATGAAGTTACTTATT	c
25	1510	1530	1550	
	TACCCTGAACTTTGAAGC	CCAGCAACCCACACCAACCA	ACTTCAGCCTCCCCATCCCTAA	Ċ
30	1570	1590	1610	
	AGCCACAGAAATAATTTA	TTCAGAAGTAAAAAAGCAG	RAATGAAACCTGAAAAAAAAAAAAA	A
	1630			
35			•	
40				
45				

(3)

5	10	30	50	
	CAGCCSTGCTCGAAGCGTT	CCTGGAGCCCAAGCTCTCC	CCACAGGTGAAGACAGGG	CCA
10	70	90	110	
	GCAGGAGACACCATGGGGC MetGlyH	ACCTCTCAGCCCCACTTCA isLeuSerAlaProLeuHi		
15	130	150	170	
20	GGGCTTCTGCTCACAGCCT GlyLeuLeuLeuThrAlas	CACTTCTAACCTTCTGGAA erLeuLeuThrFheTrpAs	CCGCCCACCACTGCCCAG nProProThrThrAlaGln	CTC Leu
	190	210	230	
25	ACTACTGAATCCATGCCAT ThrThrGluSerMetProP	TCAATGTTGCAGAGGGGAA heasnvalalaGluGlyLy	GGAGGTTCTTCTCCTTGTC sGluValLeuLeuLeuVal	CAC His
	250	270	290	
30	AATCTGCCCCAGCAACTTT AsnLeuProSinGinLeuP	TTGGCTACAGCTGGTACAA heGlyTyrSerTrpTyrLy	AGGGGAAAGAGTGGATGGG sGlyGluArgVslAspGl <sub>3</sub>	IAAC /Ash
	313	330	350	
35	CGTCAAATTGTAGGATATC ArgGln1leValGlyTyrA	CAATAGGAACTCAACAAGG NaileGlyThrGlnGlnAl	TACCCCAGGGCCCGCAAA( TACCCCAGGGCCCGCAAA( ThrProGlyProAlaas:	IAGC nSet
40	370	390	410	
40		CCAATGCATCCCTGCTGAT ProAsnAlaSerLeuLeuI		
45	430	450	470	
	ACAGGATTCTACACCCTAG ThrGlyPheTyrThrLeu	CAAGTCATAAAGTCAGATC GlnVallleLysSerAspL	CTGTGAATGAAGAAGCAAC euValasnGluGlualath	TCG;

	1030	1050	1070		
5			CTCACCTGGGGCCATTGCTG uSerProGlyAlaIleAlaG		
	1090	1110	1130		
10			AGTAGCCCTGGCATGTTTTC aValalaLeuAlaCysPheL		
	1150	1170	1190		
15		AGCTCAGGACCACTCCAATC SerSerGlyProLeuGln	CCCACCTAACAAGATGAAT	IGA	
	1210	1230	1250		
20	AGTTACTTATTCTACCCTGAACTTTGAAGCCCAGCAACCAAC				
	1270	1290	1310		
25	CCCATCCCTARCAGCCACAGAAATAATTTATTCAGAAGTAAAAAAGCAGTAATGAAACCT				
	1330				
30	Gaaarararahaaa	₩ <sub>.</sub>			
35					
			•		
40					
45					
50					

ļ

(4)

5	1	acagcacagctgacagccgtactcaggaagcttctggatcctaggcttatctccacagag	60
	51	gagaacacacagcagcagagaccatggggcccctctcagcccctccct	120
10	121	atcacttggaaggggtcctgctcacagcatcacttttaaacttctggaatccgcccaca. IleThrTrpLysGlyValLeuLeuThrAlaSerLeuLeuAsnPheTrpAsnProProThr	1.80
	181	actcccaactcaccattgaaccccagccacccaaactttctgagggcaaccatctt ThrAlaGlnValThrIleGluAlaGlnProProLysValSerGluGlyLysAspValLeu	240
15	241	ctacttgtccacaatttgccccagaatcttgctggctacatttggtacaaagggcaaatg LeuLeuValHisAsnLeuProGlnAsnLeuAlaGlyTyrIleTrpTyrLysGlyGlnMet	300
13	301	acatacgtctaccattacattacatcatatgtagtagacggtcaaagaattatatatggg ThrTyrValTyrHisTyrIleThrSerTyrValValAspGlyGlnArgIleIleTyrGly	360
	361	cctgcatacagtggaagaaagagtatattccaatgcatccctgctgatccagaatgtc ProAlaTyrSerGlyArgGluArgValTyrSerAsnAlaSerLeuLeuIleGlnAsnVal	420
20	[421	acgcaggaggatgcaggatcctacaccttacacatcataaagcgacgcgatgggactgga ThrGlnGluAspAlaGlySerTyrThrLeuHisIleIleLysArgArgAspGlyThrGly	480
	481	ggagtaactggacatttcaccttcaccttacacctggagactcccaagccctccatctcc GlyValThrGlyHis?heThrPheThrLeuHisLeuGluThrProLysProSerIleSer	540
25	541	agcagcaacttaaatcccagggaggccatggaggctgtgatcttaacctgtgatcctgcg SerSerAsnLeuAsnProArgGluAlaMetGluAlaValIleLeuThrCysAspProAla	600
	501	actccagccgcaagctaccagtggtggatgaatggtcagagcctccctatgactcacagg Thr?roAlaAlaSerTyrGlnTrpTrpMetAsnGlyGlnSerLeuProMetThrHisArg	660
30	551	ttgcagctgtccaaaaccaacaggaccctctttatatttggtgtcacaaagtatattgca LeuGlnLeuSerLysThrAsnArgThrLeuPheIlePheGlyValThrLysTyrIleAla	720
	721	gçaccctatgaatgtgaaatacgçaacccagtgagtgccagccgcagtgacccagtcacc GlyProTyrGluCysGluIleArgAsnProValSerAlaSerArgSerAspProValThr	780
35	781	ctgaatctcctcccaaagctgtccaagccctacatcacaatcaacaacttaaaccccaga LeuAsnLeuLeuProLysLeuSerLysProTyrIleThrIleAsnAsnLeuAsnProArg	840
	841	Gaçaataaggatçtettaacetteacetgtgaacetaagagtgagaactacacetacatt GluAsnLysAspValLeuThrPheThrCysGluProLysSerGluAsnTyrThrTyrIle	900
40	901	tggtggctaaatggtcagagcctccctgtcagtcccagggtaaagcgacccattgaaaac TrpT:pLeuAsnGlyGlnSerLeuProValSerProArgValLysArgProIleGluAsn	960
	961	aggatecteattetacecaatgteacgagaaatgaaacaggacettateaatgtgaaata ArgileLeuIleLeuProAsnValThrArgAsnGluThrGlyProTyrGlnCysGluIle	1020
45	1021	C999accgatatggtggcatccgcagtgacccagtcaccctgaatgtcctctatggtcca ArgAspArgTyrGlyGlyIleArgSerAspProValThrLeuAspValTeuTyrGlyPro	1080

50

	1081	<pre>gacctccccagcatttacccttcattcacctattaccgttcaggagaaaacctctacttt AspLeuProSerIleTyrProSerPheThrTyrTyrArgSerGlyGluAsnLeuTyrPhe</pre>	1140
5	1141	tcctgcttcggtgagtctaacccacgggcacaatattcttggacaattaatgggaagttt SerCysPheGlyGluSerAsnProArgAlaGlnTyrSerTrpThrIleAsnGlyLysPhe	1200
	1261	cagctatcaggacaaaagctctctatcccccaaataactacaaagcatagtgggctctat GlnLeuSerGlyGlnLysLeuSerIleProGlnIleThrThrLysHisSerGlyLeuTyr	1260
10	1261	gcttgctctctctcgtaactcagccactggcaaggaaagctccaaatccatcacagtcaaa AlaCysSerValArgAsnSerAlaThrGlyLysGluSerSerLysSerIleThrValLys	1320
	1321	gtctctgactggatattaccctgaattctactagttcctccaattccattttctcccatg ValSerAspTrpileLeuProEnd	1350
15	1381 1441 1501	gaztczcgaagagcaagaccczctctgttccagazgccctataatctggaggtggacaac tcgatgtaaatttcatgggaaaacccttgtacctgacatgtgagccactcagazctcacc	1440 1500
	1561 1621 1681 1741	aaaatgttcgacaccataacaacagctactcaaactgtaaaccaggataagaagttgatg acttcacactgtggacagtttttcaaagatgtcataacaagactccccatcatgacaagg ctccaccctctactgtctgctcatgcctgctctttcacttggcaggataatgcagtcat tagaatttcacatgtagtagcttctgagggtaacaacagagtgtcagatatgtcatctca	1560 1620 1680 1740
20	1801 1861 1921 1981	acctcaaacttttacgtaacatctcagggaaatgtggctctccatcttgcatacagggctcccaatagaaatgaacacagagatattgcctgtgtgtttgcagaagaagatggtttctatacagagataggatattgcctgtgtgtttgcagaagaagatggttctatacagagataggatacagtagaagtaaatgcacattgtgtggatggctctcaccatttcctaagagatacagtgtaaaaacgtgacagtaatactgattctagcagaataaacatgtaccacatttgcaaaaaa	1800 1860 1920 1980 2010

and

5	. 1	gggtggatcctaggctcatctccataggggagaacacacatacagcagagaccatggga MetGly	59
	50	ccctctcagccctccctgcactcagcacatcacctggaaggggctcctgctcacagca ProLeuSerAlaProProCysThrGlnHisIleThrTrpLysGlyLeuLeuThrAla	119
10	120	tcacttttaaacttctggaacctgcccaccactgcccaagtaataattgaagcccagcca SerLeuLeuAsnPheTrpAsnLeuProThrThrAlaGlnValIleIleGluAlaGlnPro	179
	180	cccaaagtttctgaggggaaggatgttcttctacttgtccacaatttgccccagaatctt ProLysValSerGluGlyLysAspValLeuLeuLeuValHisAsnLeuProGlnAsnLeu	239
15	240	actggctacatctggtacaaagggcaaatgacggacctctaccattacattacatcatat ThrGlyTyrIleTrpTyrLysGlyGlnMetThrAspLeuTyrHisTyrIleThrSerTyr	299
	300	gtagtagacggtcaaattatatatgggcctgcctacagtggacgagaaacagtatattcc ValValAspGlyGlnIleIleTyrGlyProAlaTyrSerGlyArgGluThrValTyrSer	359
20	360	aatgcatccctgctgatccagaatgtcacacaggaggatgcaggatcctacaccttacac AsnAlaSerLeuleuIleGlnAsnValThrGlnGluAspAlaGlySerTyrThrLeuHis	419
	420	atcataaagcgaggcgatgggactggaggagtaactggatatttcactgtcaccttatac IleIleLysArcGlyAspGlyThrGlyGlyValThrGlyTyrPheThrValThrLeuTyr	479
25	480	tcggagactcccaagcgctccatctccagcagcaacttaaaccccagggaggtcatggag SerGluThrPrcLysArgSerIleSerSerSerAsnLeuAsnProArgGluValMetGlu	539
	540	gctgtgcgcttaatctgtgatcctgagactccggatgcaagctacctgtggttgctgaat AlaValArgLeuIleCysAspProGluThrProAspAlaSerTyrLeuTrpLeuLeuAsn	599
30	600	ggtcagaacctccctatgactcacaggttgcagctgtccaaaaccaacaggaccctctat GlyGlnAsnLeuProMetThrHisArgLeuGlnLeuSerLysThrAsnArgThrLeuTyr	659
	660	ctatttggtgtcacaaagtatattgcagggccctatgaatgtgaaatacggaggggagtg LeuPheGlyValThrLysTyrIleAlaGlyProTyrGluCysGluIleArgArgGlyVal	719
35	720	agtgccagccgcagtgacccagtcaccctgaatctcctcccgaagctgcccatgccttac SerAlaSerArgSerAspProValThrLeuAsnLeuLeuProLysLeuProHetProTyr	779
	780	atcaccatcaacaacttaaaccccagggagaagaaggatgtgttagccttcacctgtgaa IleThrIleAsnAsnLeuAsnProArgGluLysLysAspValLeuAlaPheThrCysGlu	839
40	840	cctaagagtcggaactacacctacatttggtggctaaatggtcagagcctcccggtcagt ProLysSerArgAsnTyrThrTyrIleTrpTrpLeuAsnGlyGlnSerLeuProValSer	899
	900	ccgagggtaaagcgacccattgaaaacaggatactcattctacccagtgtcacgagaaat ProArgValLysArgProIleGluAsnArgIleLeuIleLeuProSerValThrArgAsn	959
45	960	gaaacaggaccctatcaatgtgaaatacgggaccgatatggtggcatccgcagtaaccca GluThrGlyProTyrGlnCysGluIleArgAspArgTyrGlyGlyIleArgSerAsnPro	1019

1020	gtcaccctgaatctcctctatggtccagacctccccagaatttacccttacttcacctat valThrLeuAsnValLeuTyrGlyProAspLeuProArgIleTyrProTyrPheThrTyr	1079
1080	taccgttcaggagaaaacctcgacttgtcctgctttgcggactctaacccaccggcagag TyrArgSerGlyGluAsnLeuAspLeuSerCysPheAlaAspSerAsnProProAlaGlu	1139
1140	tatttttggacaattaatgggaagtttcagctatcaggacaaaagctctttatcccccaa TyrPheTrpTh:IleAsnGlyLysPheGlnLeuSerGlyGlnLysLeuPheIleProGln	1199
1200	attactacaaatcatagcgggctctatgcttgctctgttcgtaactcagccactggcaag IleThrThrAsnHisSerGlyLeuTyrAlaCysSerValArgAsnSerAlaThrGlyLys	1259
1260	gaaatctccaaatccatgatagtcaaagtctctggtccctgccatggaaaccagacaga	1319
1320	tctcattaatggctgccacaatagagacactgagaaaaagaacaggttgataccttcatg SerHisEnd	1379
1380 1440 1500 1560	aaattcaagacaaagaagaaaaaggctcaatgttattggactaaataatcaaaaggataa tgttttcataatttttattggaaaatgtgctgattcttggaatgttttattctccagatt tatgaacttttttttcttcagcaattggtaaagtatacttttgtaaacaaaaattgaaaca tttgcttttqctctctatctgagtgcccccc 1591	1439 1499 1559
	1080 1140 1200 1260 1320 1380 1440 1500	ValThrLeuAsnValLeuTyrGlyProAspLeuProArgIleTyrProTyrPheThrTyr  1080 taccgttcagcagaaaacctcgacttgtcctgctttgcggactctaacccaccggcagag TyrArgSerGlyGluAsnLeuAspLeuSerCysPheAlaAspSerAsnProProAlaGlu  1140 tattttggacaattaatgggaagtttcagctatcaggacaaaagctctttatccccaa TyrPheTrpThrIleAsnGlyLysPheGlnLeuSerGlyGlnLysLeuPheIleProGln  1200 attactacaaatcatagcgggctctatgcttgctctgttcgtaactcagccactggcaag IleThrThrAsnHisSerGlyLeuTyrAlaCysSerValArgAsnSerAlaThrGlyLys  1260 gaaatctccaaatccatgatagtcaaagtctctggtccctgccatggaaaccagacaga

20

- 2. Replizierbares rekombinantes Kloniervehikel mit einem eine Nucleinsäure nach Anspruch 1 umfassenden Insert.
- Zelle, die mit einem rekombinanten Kloniervehikel nach Anspruch 2 transfiziert, infiziert oder injiziert ist.
  - 4. Verfahren zur Herstellung eines Polypeptids, umfassend die Schritte
    - (a) des Kultivierens der Zelle nach Anspruch 3,
    - (b) des Gewinnens des durch diese Zelle exprimierten Polypeptids.

30

- 5. Verfahren zur Herstellung eines gegen ein Polypeptid gerichteten Antikörpers, umfassend die Schritte
  - (a) des Herstellens des Polypeptids durch das Verfahren des Anspruchs 4,
  - (b) des Injizierens des Polypeptids in einen Wirt, der zur Bildung von Antikörpern befähigt ist, und
  - (c) des Gewinnens der Antikörper.

35

### Revendications

1. Acide nucléique comprenant une séquence de bases qui code pour une séquence peptidique, caractérisé en ce que le groupe d'acides nucléiques est de l'ADN choisi parmi le groupe de cinq séquences ci-après :

45

40

	10	30	50	•
	CAGCCGTGCTCGAAGCGTT	CCTGGAGCCCAAGCTCTCCT	CCACAGGTGAAGACAGG	JCCA
5				
	70	90	110	
10		ACCTCTCAGCCCCACTTCAC isLeuSerAlaProLeuHis		
	130	150	170	
15	GGGCTTCTGCTCACAGCCT GlybeubeubeuThrAleS	CACTTCTAACCTTCTGGAAC SecteuleuThtPheTtpAsr	CCGCCCACCACTGCCCA NProProThrhIALaG1	GCTC nLst
	190	210	230	
20		FTCAATGTTGCAGAGGGGAA( PheAsnValAlaGluGlyLy:		
	250	270	290	
25		TTTGGCTACAGCTGGTACAAA TheGlyTyrSerTrpTyrLys		
	. 310	330	350 ·	
30	CGTCAAATTGTAGGATATG ArgGlnIleValGlyTyr/	CCAATAGGAACTCAACAAGC AlaileGlyTh:GlnGlnAla	TACCECAGSGCCCGCAAA aThcPcoGlyPcoAlaAs	.CAGC .nSer
35	370	390	410 .	
,,,		CCCAATGCATCCCTGCTGAT ProAsnAlaSerieuLeuIl		
40	130	450	470	
		CAAGTCATAAAGTCAGATCT GlnVallletysSerAspLe		
45				

	493	510	530	
5	CAGTTCCATGTATACCC GlaPheHisValTy:P:	CGGAGCTGCCCAAGCCCTCCATC CDGluLeuPcoLys?coSecīle	TCCAGCAACAACTCCAACCCT SerSerAsnAsnSerAsnPro	
	5 5 0	570	590	
10	GTGGAGGACAAGGATGG ValGluAsplysAspAl	CTGTGGCCTTCACCTGTGAACCT lavalalaPheThrCysGluPro	GAGACTCAGGACACAACCTAC GluthrGlnAspThrThrTy:	
	510	630	650	
15	CTGTGGTGGATAAACA LeuTrpTrpIleAsnA	ATCAGAGCCTCCCGGTCAGTCCC snGlnSerLeuProValSerPro	AGGCTGCAGCTGTCCAATGGC ArgLeuGlnLeuSerAsnGly	· - •
	<b>57</b> 0	590	710	•
20		TACTCAGTGTCACAAGGAATGA euleuservalthrargasnas		
	732	.750	770	
25		GTGCGAACCGCAGTGACCCAGT eralaasnargSerAspProVa		
30	793	810	830	
		TTTCCCCTTCAGACACCTATTA leserProserAspThrTyrTy		
35	850	870	890	
		CCTCTAACCCACCTGCACAGTA AlaSerAsnProProAlaGlnTy		
40	910	930	950	
		CAAGAGCTCTTTATCCCTAACAT SinGluLeuPheIleProAsnIl		
45	975	990	,1010	
	TATACCTGCCACGCCA TyrThrCysHisAla	AATAACTCAGTCACTGGCTGCA AsnasnSecValThcGlyCysa	ACAGGACCACAGTCAAGACGA snAcgThcThcValLy.sThcI	; c

	1030	1050	1070
5			TCACCTGGGGCCATTGCTGGC SerProGlyAlaIleAlaGly
	1090	1110	1130
10			GTAGCCCTGGCATGTTTTCTC
	1150	1170	1190
15			CTCACAGAGCACAAACCCTCA bLeuThrGluHisLysProSer
	1210	1230	1250
20			TAACAAGATGAATGAAGTTACT DASnLysMetAsnGluValTh:
25	1275	1290	1310
			ACCAACTTCAGCCTCCCCATCC nProThrSerAlaSerProSer
30	1333	1350	1370
		AATTTATTCAGAAGTAAAAA eiletyrSerGluVallysLy	GCAGTAATGAAACCTGTCCTGC sGln
35	1390	1410	1430
	TCACTGCAGTGCTGATG	TATTTCAAGTCTCTCACCCTC	ATCACTAGGAGATTCCTTTCCC
40	1450	1470	1490
	CTGTAGGGTAGAGGGGT	GGGGACAGAACAACTTTCTC	CTACTCTTCCTTCCTAATAGGC
45	1510	1530	1550
	ATCTCCAGGCTGCCTGG	TCACTGCCCCTCTCTCAGTGT	CAATAGATGAAAGTACATTGGG
50	1570	1590	1610
	AGTCTGTAGGAAACCCA	ACCTTCTTGTCATTGAAATT	rggcaaagctgacttrgggaaag

	1533	1650	1670	
5	AGGGACCAGAACTTCCCCT	CCCTTCCCCTTTTCCCAAC	CTGGACTTGTTTTAAACTTS	C C
J	1690	1710	1730	
	TGTTCAGAGCACTCATTCC	TTCCCACCCCAGTCCTGT	CCTATCACTCTAATTCGGAT	TT
10	1750	1770	1790	
	GCCATAGCCTTGAGGTTAT	GTCCTTTTCCATTAAGTAC	ATGTGCCAGGAAACAGCGAG	AG
15	1810	1830	1850	
	AGAGAAAGTAAACGGCAGT	AATGCTTCTCCTATTTCTC	CAAAGCCTTGTGTGAACTAG	CA
20	1870	1890	1910	
	AAGAGAAGAAAATCAAATA	TATAACCAATAGTGAAATG	CCACAGGTTTGTCCACTGTC	AG
25	1930	1950	1970	
	GGTTGTCTACCTGTAGGAT	CAGGGTCTAAGCACCTTGG	TGCTTAGCTAGAATACCACC	TÀ
30	1990	2010	2030	
	ATCCTTCTGCAAGCCTGT	CCTTCAGAGAACCCACTAGA	AGCAACTAGGAAAAATCACT	``FG
35	2050	2070	2090	
	CCAAAATCCAAGGCAATTI	CCTGATGGAAAATGCAAAAG	CACATATATGTTTTAATATC	:TT
40	2110	2130	2150	
	TATGGGCTCTGTTCAAGG	CAGTGCTGAGAGGGAGGGG	PTATAGCTTCAGGAGGGAACC	IAG
45	2170	2190	2210	
	CTTCTGATAAAQACAATC	TGCTAGGAACTTGGGAAAG	SAATCAGAGAGCTGCCCTTC	5GC

	2230	2250	2 2.7 0	
	CATTATTTAAATTĞTTAA	NGAATACACAATTTGGGGT/	ATTGGGATTTTTCTCCTTTTCT	Ċ
5				
	2293	2310	2330	
	TGAGACATTICACCATTT	TAATTTTTGTAACTGCTȚA	TTTATGTGAAAAGGGTTATTT	T
10	2350	2370	2390	
			 GGTGAAAGAAACCACCGAAATC	
15	ACTINGCTIAGETATOTE	AGCEMICCONTIOCCTIA	oo i okka daa caa caa caa caa caa caa caa caa ca	. •
,,	2410	2430	2450	
	CTCAGGTCCCTTGGTCAG	GAGCCTCTCAAGATTTTTT	TTGTCAGAGGCTCCAAATAGAA	À
20	2422		. 2510	
	2470	2490	2510-	
	ATAAGAAAAGGTTTTCTT	CATTCATGGCTAGAGCTAG	ATTTAACTCAGTTTCTAGGCA	I C
25	2530	2550	2570	
	TCAGACCAATCATCAACT	ACCATTCTATTCCATGTTT	GCACCTGTGCATTTTCTGTTT	S.C
30	2590	2610	2630	
	CCCCATTCACTTTGTCAG	GAAACCTTGGCCTCTGCTA	AGGTGTATTTGGTCCTTGAGA	ÀĞ
35	2650	2670	259C	
	TGGGAGCACCCTACAGGG	ACACTATCACTCATGCTGC	STGGCATTGTTTACAGCTAGAA	A G
40	2710	2730	2750	
	CTGCACTGGTGCTAATGC	CCCTTGGGAAATGGGGCTC	GTGAGGAGGAGGATTATAACTT	ΑG
45	2770	2790	2010	
	GCCTAGCCTCTTTTAACA	AGCCTCTGAAATTTATCTT	PTCTTCTATGGGGTCTATAAAT	CT
50	2830	2850	2870	
	ATCTTATAATAAAAAGGA	AAGGACAGGAGG <mark>AAGACAG</mark>	GCAAATGTACTTCTCACCCAGT	CT

	2890	2910	2930
	TCTACACAGATGGAATCT	CTTTGGGGCTAAGAGAAAGG	FTTTATTCTATATTGCTTACCT
5	2950	2970	2990
	•		
10	- GAICICAIGIIAGGCCIA	ACAGCTTTCTCCAGGAGA	ITAGCTTGGAGTTCTCTATACT ·
	3010	3030	3050
	CAGGTACCTCTTTCAGGG	TTTTCTAACCCTGACACGGA	CTGTGCATACTTTCCCTCATCC
15	3070	3090	3330
	•		3110
20	AIGCIGIGIGIAII	TAATTITICCIGGCTAAGAT	CATGTCTGAATTATGTATGAA
	3130	3150	3170
	ATTATTCTATGTTTTTAT	'AATAAAAATAATATATCAGA	CATCGAAAAAAAAA,
25	·		
	•		
30			
00			
35			
40			
70			
45			
<b>5</b> 0			
50			

(2)

	437	450	470	
_	ACAGGATTCTACACCCTAC ThrGlyPheTyrThrLeuG	RAGTCATAAAGTCAGATCT lnValileLysSerAspLev	rGTGAATGAAGAAGCAACT uValAsnGluGluAlaThr	SGA Gly
5	490	510	 530	
10	CAGTTCCATGTATACCCGG	AGCTGCCCAAGCCCTCCAT( luLeuProLysProSerIl)		
	550	570	590	
15	GTGGAGGACAAGGATGCTG ValGluAspLysAspAlaV	TGGCCTTCACCTGTGAACC	TGAGACTCAGGACACAACC oGluThrGlnAspThrThr	TAC Tyr
	610	630	650	
20		AGAGCCTCCCGGTCAGTCC lnSerLeuProValSerPr		
05	670	690	710	
25		TCAGTGTCACAAGGAATGA euSerValThrArgAsnAs		
30	730	750	770	
		CGAACCGCAGTGACCCAGT LlaAsnArgSerAspProVa		
35	. 790	810	830	
		CCCCTTCAGACACCTATTA erProSerAspThrTyrTy	- • -	
40				
45				

	550	6 / 0	690
5	CTCTCCTSCTATGCAGCCT LeuSerCysTyrAlaAlaS	OATBACCCACCTGCACAGTAC Secambrosforfication	TCCTGGCTTATCAATGGAACA SerTrpLeuIleAsnGlyThr
	910	930	950
10	TTCCAGCARAGCACACAAC PheGlnGlnSerThrGlnC	GAGCTCTTTATCCCTAACATC	ACTGTGAATAATAGTGGATCO ThrValAsnAsnSerGlySer
	970	990	1010
15	TATACCTGCCACGCCAATA TyrThrCysHisAlaAsna	AACTCAGTCACTGGCTGCAAC AsnSerValThrGlyCysAsr	AGGACCACAGTCAAGACGATCAAGAACAAGACAAGACAAGACAAGACAAGACAAGACAAGAACAAGAACAAGAACAAAAAA
	1030	1050	1070
20			ATCAAAGCCAGCAAGACCACA hileLysAlaSerLysThrTh
25	1090	1110	1130
			CACAAATGACACTGGAATCTC
10	- 1150	1170	1190
			GGAGAGGATGAAGCTGTCCCA rGluArgMetLysLeuSerGl
5	1210	1230	1250
			GGATGCTGGGACGTATTGGTG uAspAlaGlyThrTyrTrpCy
0			
5			

	1270	1290	1310
5	GAGGTCTTCAACCCAAT GluValPheAsnProIl	CAGTAAGAACCAAAGCGACCC eSerLysAsnGlnSerAspPr	CATCATGCTGAACGTAAACTAT oilemetLeuAsnValAsnTyr
	1330	1350	1370
10	AATGCTCTACCACAAGA AsnalaLeuFroGlnGl	AAATGGCCTCTCACCTGGGGC UASnGlyLeuSerProGlyAl	CATTGCTGGCATTGTGATTGGA alleAlaGlyIleValIleGly
	1390	1410	1430
15	GTAGTGGCCCTGGTTGC ValValAlaLeuValAla	TCTGATAGCAGTAGCCCTGGC eLeuIleAlaValAlaLeuAl	ATGTTTTCTGCATTTCGGGAAG aCysPheLeuHisPheGlyLys
20	1450	1470	1490
	ACCGGCAGCTCAGGACCZ ThrGlySerSerGlyPro	ACTCCAATGACCCACCTAACA DLeuGln	AGATGAATGAAGTTACTTATTC
25	1510	1530	1550
	TACCCTGAACTTTGAAG	CCCAGCAACCCACACCAA	CTTCAGCCTCCCCATCCCTAAC
30	1570	1590	1610
	AGCCACAGAAATAATTT	ATTCAGAAGTAAAAAGCAGT	AATGAAACCTGAAAAAAAAA
35	1630		
	*****		
40			
45			

(3)

7.0

15

25

30

35

40

45

50

10 30 50

CAGECSTGETESAAGESTTEETSGAGCECAAGETETECTCCACAGGTSAAGACAGGGCCA
10

GCAGGAGACACCATGGGGCACCTCTCAGCCCCACTTCACAGAGTGCGTGTACCCTGGCAG
MetGlyHisLeuSerAlaProLeuHisArgValArgValProTrpGln

110

90

130 150 170

GGGCTTCTGCTCACAGCCTCACTTCTAACCTTCTGGAACCCGCCCACCACTGCCCAGCTCGlyLeuLeuThrAlaSerLeuLeuThrPheTrpAsnProProThrThrAlaGinLeu

190 210 230

ACTACTGAATCCATGCCATTCAATGTTGCAGAGGGGGAAGGAGGTTCTTCTCCTTGTCCAC ThrThrGluSerMetProPheAsnValAlaGluGlyLysGluValLeuLeuLeuValHis

250 270 290

AATCTGCCCAGCAACTTTTTGGCTACAGCTGGTACAAAGGGGAAAGAGTGGATGGCAACASnLeuProGlnGlnLeuPheGlyTyrSerTrpTyrLysGlyGluArgValAspGlyAsn

313 330 350

CGTCAAATTGTAGGATATGCAATAGGAACTCAACAAGCTACCCCAGGGCCCGCAAACAGC ArgGlnIleValGlyTyrAlaIleGlyThrGlnGlnAlaThrProGlyProAlaAsnSer

370 390 410

GGTCGAGAGACATATACCCCAATGCATCCCTGCTGATCCAGAACGTCACCCAGAATGACGlyArgGluThrIleTyrProAsnAlaSerLeuLeuIleGlnAsnValThrGlnAsnAsp

430 450 470

	493	510	530
5	CAGTTCCATGTATACCCCGGInPheHisValTy:2:0	GGAGCTGCCCAAGCCCTCCATC oGluLeuProLysProSerIle	TCCAGCAACAACTCCAACCCT SerSerAsnAsnSerAsnPro
	550	570	590
10		TGTGGCCTTCACCTGTGAACCT aValAlaPheThrCysGluPro	
	610	630	650
15			AGGCTGCAGCTGTCCAATGGC ArgLeuGlnLeuSerAsnGly
	670	690	710
20			ACAGGACCCTATGAGTGTGAA oThrGlyProTyrGluCysGlu
25	730	750	770
30			TACCTTGAATGTCACCTATGGC LThrLeuAsnValThrTyrGly
	797	810	830
35		TTCCCCTTCAGACACCTATTAC eSerProSerAspThrTyrTyr	
	8.5.3	870	. 890
40		CTCTAACCCACCTGCACAGTAC aSerAsnProProAlaGlnTyr	
	910	930	950
45		AGAGCTCTTTATCCCTAACATC	
	970	990	1010
50	TATACCTGCCACGCCAA TycThcCysHisalaAs	TAACTCAGTCACTGGCTGCAAC	:AGGACCACAGTCAAGACGATC :ArgThrThrVallysThrIle

	1030	1050	,	1070	
5	ATAGTCACTGATAATGCTGT. TlevalThraspasnalaLe	ACCACAAGAAAATO uzcoGlnGluAsnO	GCCTCTC.	CCTGGGGCCATTG ProGlyAlaIleA	crosc lagly
	1090	1110		1130	
10	ATTGTGATTGGAGTAGTGGC ilevalileGlyvalValAl	CCTGGTTGCTCTG/ aLeuValAlaLeu	ATAGCAGT: IleAlaVa	AGCCCTGGCATGTT IAlaLeuAlaCysE	TTCTS PheLeu
	1150	1170		1190	
15	CATTTCGGGAAGACCGGCAG HisPheGlyLysThrGlySe	CTCAGGACCACTC rserGlyProLeu	CAATGACC Gln	CACCTAACAAGATO	GAATGA
20	1210	1230		1250-	
	AGTTACTTATTCTACCCTGA	ACTTTGAAGCCCA	GCAACCCA	CACAACCAACTTC	AGCCTC
25	1270	1290		1310	
	CCCATCCCTAACAGCCACAC	SAAATAATTTATTC	AGAAGTAA	AAAAGCAGTAATG	AAACCT
30	1330				
	GAAAAAAAAAAAAAAA				
35					
40					
45					

(4)

	:	acageacagetgacageegtacteaggaagettetggateetaggettateteeacagag	60
5	51	gagaacacacagcagcagagaccatggggcccctctcagcccctccct	120
	121	atcacttggazggggtcctgctcacagcatcacttttaaacttctggaatccgcccaca IleThrTrpLysGlyValLeuLeuThrAlaSerLeuLeuAsnPheTrpAsnProProThr	190
10	181	actgcccaagtcacgattgaagcccagccacccaaagtttctgaggggaaggatgttctt ThrAlaGlnValThrIleGluAlaGlnProProLysValSerGluGlyLysAspValLeu	240
	241	ctacttgtccacaatttgccccagaatcttgctggctacatttggtacaaagggcaaatg LeuLeuValHisAsnLeuProGlnAsnLeuAlaGlyTyrIleTrpTyrLysGlyGlnMet	300
15	301	acatacgtctaccattacattacatcatatgtagtagacgqtcaaagaattatatggg ThrTyrValTyrHisTyrIleThrSerTyrValValAspGlyGlnArgIleIleTyrGly	360
	361	cctgcatacaçtçgaagaaagagtatattccaatgcatccctgctghtccagaatçtc ProAlaTyrSerGlyArgGluArgValTyrSerAsnAlaSerLeuLeuIleGlnAsnVal	420
20	.421	acgcaggaggatgcaggatcctacaccttacacatcataaagcgacgcgatgggactgga ThrGlnGluAspAlaGlySerTyrThrLeuHisIleIleLysArgArgAspGlyThrGly	480
	481	ggagtaactggacatttcaccttcaccttacacctggagactcccaagccctccatctcc GlyValThrGlyHisPheThrPheThrLeuHisLeuGluThrProLysProSerIleSer	540
25	541	agcagcaacttaaatcccaggcaggccatggaggctgtgatcttaacctgtgatcctgcg SerSerAsnLeuAsnProArgGluAlaMetGluAlaValIleLeuThrCysAspProAla	600
	501	actccagccgcaagctaccagtggtggatgaatggtcagagcctccctatgactcacagg Thr?roAlaAlaSerTyrGlnTrpTrpMetAsnGlyGlnSerLeuProMetThrHisArg	660
<i>30</i>	PP 1	ttgcagctgtccaaaaccaacaggaccctctttatatttggtgtcacaaagtatattgca LeuGliLeuSerLysThrAsnArgThrLeuPheIlePheGlyValThrLysTyrIleAla	720
	721	ggaccctatgaatgtgaaatacggaacccagtgagtgccagccgcagtgacccagtcacc GlyProTyrGluCysGluIleArgAsnProValSerAlaSerArgSerAspProValThr	780
35	731	ctgaatctcctcccaaagctgtccaagccctacatcacaatcaacaacttaaaccccaga LeuAsnLeuLeuProLysLeuSerLysProTyrIleThrIleAsnAsnLeuAsnProArg	840
	841	<pre>gagaataaggatgtcttaaccttcacctgtgaacctaagagtgagaactacacctacatt GluAsnLysAspValLeuThrPheThrCysGluProLysSerGluAsnTyrThrTyrIle</pre>	900
40	901	tggtggctaaatggtcagagcctccctgtcagtcccagggtaaagcgacccattgaaaac TrpTrpLeuAsnGlyGlnSerLeuProValSerProArgValLysArgProIleGluAsn	960
	961	aggatecteattetacecaatgteacgagaaatgaaacaggacettateaatgtgaaata ArgileLeuIleLeuProAsnValThrArgAsnGluThrGlyProTyrGlnCysGluIle	1020
45	1021	cgggaccgatatggtggcatccgcagtgacccagtcaccctgaatgtcctctatggtcca ArgaspArgTyrGlyGlyIleArgSerAspProValThrLevAspValLevTyrGlyPro	1080

	1081	gacttccccagcatttacccttcattcacctattaccgttcaggagaaaacctctacttt AspleuProSerileTyrProSerPh,ThrTyrTyrArgSerGlyGluAsnLeuTyrPhe	1140
5	1141	tcctqcttcqqtqagtctaacccacgggcacaatattcttggacaattaatgggaagttt SerCysPheGlyGluSerAsnProArgAlaGlnTyrSerTrpThrIleAsnGlyLysPhe	1200
	1201	cagotatoagçacaaaagotototatococcaaataactacaaagoatagtgggototat GlnLeuSerGlyGlnLysLeuSerIleProGlnIleThrThrLysHisSerGlyLeuTyr	1263
10	1251	gcttgctctgttcgtaactcagccactggcaaggaaagctccaaatccatcacagtcaaa AlaCysSerValArgAsnSerAlaThrGlyLysGluSerSerLysSerIleThrValLys	1329
	1321	gtctctgactggatattaccctgaattctactagttcctccaattccattttctcccatg ValSerAspTrpIleLeuProEnd	1380
15	1381	gaatcacgaagagcaagacccactctgttccagaagccctataatctggaggtggacaac	1440
	1441	tcgatgtaaatttcatgggaaaacccttgtacctgacatgtgagccactcagaactcacc aaaatgttcgacaccataacaacagctactcaaactgtaaaccaggataagaagttgatg	1500 1560
	1561	acttcacactgtggacagtttttcaaagatgtcataacaagactccccatcatgacaagg	1620
	1621	ctccaccetctactgtctgctcatgcctgcctctttcacttggcaggataatgcagtcat	1680
	1681	tagaatttcacatgtagtagcttctgagggtaacaacagagtgtcagatatgtcatctca	1740
	1741	acctcaaacttttacgtaacatctcagggaaatgtggctctctccatcttgcatacaggg	1800
20	1801	ctcccaatagaaatgaacacagagatattgcctgtgtgttttgcagagaagatggtttcta	1860
	1861	tazagagtaggaaagctgaaattatagtagagtctcctttaaatgcacattgtgtggatg	1920
	1921	gctctcaccatttcctaagagatacagtgtaaaaacgtgacagtaatactgattctagca	1980
	1981	gaataaacatgtaccacatttgcaaaaaa	2010

and

(5)

5	1	gggtggatcctaggctcatctccataggggagaacacacatacagcagagaccatggga MetGly	59
	<b>50</b>	cccctctcagcccctccctgcactcagcacatcacctggaaggggctcctgctcacagca ProLeuSerAlaProProCysThrGlnHisIleThrTrpLysGlyLeuLeuLeuThrAla	119
10	120	tcacttttaaacttctggaacctgcccaccactgcccaagtaataattgaagcccagcca SerLeuLeuAsn?heTrpAsnLeuProThrThrAlaGlnValIleIleGluAlaGlnPro	179
	180	cccaaagtttctgaggggaaggatgttcttctacttgtccacaatttgccccagaatctt ProLysValSerGluGlyLysAspValLeuLeuLeuValHisAsnLeuProGlnAsnLeu	239
15	240	actggctacatctggtacaaagggcaaatgacggacctctaccattacattacatcatat ThrGlyTyrIleTrpTyrLysGlyGlnMetThrAspLeuTyrHisTyrIleThrSerTyr	299
	300	gtagtagacgçtcaaattatatatgggcctgcctacagtggacgagaaacagtatattcc ValValAspGlyGlnIleIleTyrGlyProAlaTyrSerGlyArgGluThrValTyrSer	359
20	360	aatgcatccctgctgatccagaatgtcacacaggaggatgcaggatcctacaccttacac AsnAlaSerLeuLeuIleGlnAsnValThrGlnGluAspAlaGlySerTyrThrLeuHis	419
	420	atcataaagcçaggcgatgggactggaggagtaactggatatttcactgtcaccttatac IleIleLysArçGlyAspGlyThrGlyGlyValThrGlyTyrPheThrValThrLeuTyr	479
25	480	tcggagactcccaagcgctccatctccagcagcaacttaaaccccagggaggtcatggag SerGluThrProLysArgSerIleSerSerSerAsnLeuAsnProArgGluValMetGlu	539
	540	gctgtgcgcttaatctgtgatcctgagactccggatgcaagctacctgtggttgctgaat AlaValArgLeuIleCysAspProGluThrProAspAlaSerTyrLeuTrpLeuLeuAsn	599
30	600	ggtcagaacctccctatgactcacaggttgcagctgtccaaaaccaacaggaccctctat GlyGlnAsnLeuProMetThrHisArgLeuGlnLeuSerLysThrAsnArgThrLeuTyr	659
	660	ctatttggtgtcacaaagtatattgcagggccctatgaatgtgaaatacggaggggagtg LeuPheGlyValThrLysTyrIleAlaGlyProTyrGluCysGluIleArgArgGlyVal	719
35	720	agtgccagccgcagtgacccagtcaccctgaatctcctcccgaagctgcccatgccttac SerAlaSerArgSerAspProValThrLeuAsnLeuLeuProLysLeuProMetProTyr	779
	780	atcaccatcaacaacttaaaccccagggagaagaaggatgtgttagccttcacctgtgaa IleThrIleAsnAsnLeuAsnProArgGluLysLysAspValLeuAlaPheThrCysGlu	839
40	840	cctaagagtcggaactacacctacatttggtggctaaatggtcagagcctcccggtcagt ProLysSerArgAsnTyrThrTyrIleTrpTrpLeuAsnGlyGlnSerLeuProValSer	899
70	900	ccgagggtaaagcgacccattgaaaacaggatactcattctacccagtgtcacgagaaat ProArgValLysArgProIleGluAsnArgIleLeuIleLeuProSerValThrArgAsn	959
	960	gaaacaggaccctatcaatgtgaaatacgggaccgatatggtggcatccgcagtaaccca GluThrGlyProTyrGlnCysGluIleArgAspArgTyrGlyGlyIleArgSerAsnPro	1019

45

	1920	gtcaccetgaatgtcetetatggtccagacetececagaatttaccettacttcacetat ValThrLeuAsnValLeuTyrGlyProAspLeuProArgIleTyrProTyrPheThrTyr	1079
5	1080	taccgttcaggagaaacctcgacttgtcctgctttgcggactctaacccaccggcagag TyrArgSerGlyGluAsnLeuAspLeuSerCysPheAlaAspSerAsnProProAlaGlu	1139
	1140	tatttttggacaattaatgggaagtttcagctatcaggacaaaagclctttatcccccaa Tyr?heTrpThrIleAsnGlyLysPheGlnLeuSerGlyGlnLysLeuPheIleProGln	1199
10	1200	attactacaaatcatagcgggctctatgcttgctctgttcgtaactcagccactggcaag IleThrThrAsmmisSerGlyLeuTyrAlaCysSerValArgAsnSerAlaThrGlyLys	1259
	1250	gaaatctccaaatccatgatagtcaaagtctctggtccctgccatggaaaccagacaga	1319
15	1320	totoattaatggotgooacaatagagacactgagaaaaagaacaggttgatacottoatg SerHisEnd	1379
	1380 1440 1500 1560	aaattcaagacaaagaagaaaaaggctcaatgttattggactaaataatcaaaaggataa tgttttcataatttttattggaaaatgtgctgattcttggaatgtttättctccagatt tatgaactttttttttcttcagcaattggtaaagtatacttttgtaaacaaaaattgaaaca tttgcttttgctctctatctgagtgccccccc 1591	1439 1499 1559

- 2. Véhicule de clonage recombinant apte à une réplication, comportant un produit d'insertion comprenant un acide nucléique selon la revendication 1.
- 25 3. Cellule qui a été transfectée, infectée par un véhicule de clonage recombinant selon la revendication 2, ou à laquelle on a injecté ce dernier.
  - 4. Procédé pour préparer un polypeptide, ledit procédé comprenant les étapes consistant à :
    - (a) cultiver la cellule selon la revendication 3, et
    - (b) récupérer le polypeptide exprimé par ladite cellule.
  - 5. Procédé pour préparer un anticorps dirigé contre un polypeptide, ledit procédé comprenant les étapes consistant à :
    - (a) préparer ledit polypeptide par le procédé selon la revendication 4,
    - (b) injecter ledit polypeptide dans un hôte capable de produire des anticorps, et
    - (c) récupérer lesdits anticorps.

50

20

30

35

40

45

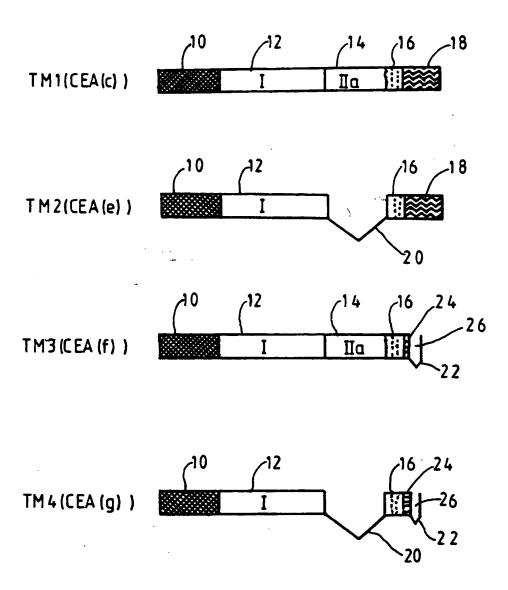


FIG.1

	CAGTTCCATGTATACCCGC GlnPheHisValTyrProC	SAGCTGCCCAAGCCCTCCAT SluLeuProLysProSerIle	CTCCAGCAACAACTCCAACCCT eSerSerAsnAsnSerAsnPro
5	550	570	590
10	GTGGAGGACAAGGATGCTC ValGluAspLysAspAla\	GTGGCCTTCACCTGTGAACC	IGAGACTCAGGACACAACCTAC DGluThrGlnAspThrThrTyr
	610	630	650
15	CTGTGGTGGATAAACAATC LeuTrpTrpIleAsnAsnC	CAGAGCCTCCCGGTCAGTCC ClnSerLeuProValSerPro	CAGGCTGCAGCTGTCCAATGGC DArgLeuGlnLeuSerAsnGly
	670	690	710
20	AACAGGACCCTCACTCTAC AsnArgThrLeuThrLeuI	TCAGTGTCACAAGGAATGA LeuSerValThrArgAsnAsi	CACAGGACCCTATGAGTGTGAA OThrGlyProTyrGluCysGlu
	· 730	750	770
25	ATACAGAACCCAGTGAGTC IleGlnAsnProValSerA	CGAACCGCAGTGACCCAGTG	CACCTTGAATGTCACCTATGGC ThrLeuAsnValThrTyrGly
	790	810	830
30	CCGGACACCCCCACCATTI ProAspThrProThrIleS	CCCCTTCAGACACCTATTAG	CCGTCCAGGGGCAAACCTCAGC ArgProGlyAlaAsnLeuSer
	850	870	890
35	CTCTCCTGCTATGCAGCCT LeuSerCysTyrAlaAlaS	CTAACCCACCTGCACAGTAG erAsnProProAlaGlnTy:	CTCCTGGCTTATCAATGGAACA SerTrpLeuileAsnGlyThr
40	910	930	950
	TTCCAGCAAAGCACACAAG PheGlnGlnSerThrGlnG	AGCTCTTTATCCCTAACATC	CACTGTGAATAATAGTGGATCC ThrValAsnAsnSerGlySer
45	970	990	1010
	TATACCTGCCACGCCAATA TyrThrCysHisAlaAsnA	ACTCAGTCACTGGCTGCAAC snSerValThrGlyCysAsr	AGGACCACAGTCAAGACGATC AArgThrThrValLysThrIle
50	1030	1050	1070

			AATCAAAGCCAGCAAGACC InIleLysAlaSerLysThr	
5	1090	1110	1130	
	GTCACAGGAGATAAGGACT	CTGTGAACCTGACCTGCT	CCACAAATGACACTGGAATC cThrAsnAspThrGlyIle	
10	1150	1170	1190	
15	ATCCGTTGGTTCTTCAAAA	ACCAGAGTCTCCCGTCCT	CGGAGAGGATGAAGCTGTCC crGluArgMetLysLeuSer	CAG Gln
	1210	1230	1250	
20	GGCAACACCACCCTCAGCA GlyAsnThrThrLeuSerI	TAAACCCTGTCAAGAGGGA leAsnProValLysArgG	AGGATGCTGGGACGTATTGG LUAspAlaGlyThrTyrTrp	TGT Cys
	1270	1290	1310	
25	GAGGTCTTCAACCCAATCA GluValPheAsnProIleS	GTAAGAACCAAAGCGACCG erLysAsnGlnSerAspP:	CCATCATGCTGAACGTAAAC COIleMetLeuAsnValAsn	TAT Tyr
	1330	1350	1370	
30			CCATTGCTGGCATTGTGATT LalleAlaGlyIleValIle	
	1390	1410	1430 .	
35	GTAGTGGCCCTGGTTGCTC ValValAlaLeuValAlaL	TGATAGCAGTAGCCCTGGG eulleAlaValAlaLeuAl	CATGTTTTCTGCATTTCGGG aCysPheLeuHisPheGly	AAG Lys
40	1450	1470	1490	
	ACCGGCAGGGCAAGCGACC ThrGlyArgAlaSerAspG	AGCGTGATCTCACAGAGCA lnArgAspLeuThrGluHi	CAAACCCTCAGTCTCCAAC .sLysProSerValSerAsn	CAC His
45	1510	1530	1550	
	ACTCAGGACCACTCCAATG ThrGlnAspHisSerAsnA	ACCCACCTAACAAGATGAA spProProAsnLysMetAs	ATGAAGTTACTTATTCTACC INGluValThrTyrSerThr	CTG Leu
50	1570	1590	1610	-
			·	-

		CCCACACAACCAACTTCAGC ProThrGlnProThrSerAl				
5	1630	1650	1670			
	GAAATAATTTATTCAGAA GluileileTyrSerGlu	GTAAAAAAGCAGTAATGAAA VallyslysGln	CCTGTCCTGCTCACTGCAG	rgc		
10						
	1690	1710	1730			
	TGATGTATTTCAAGTCTC	TCACCCTCATCACTAGGAGA	TTCCTTTCCCCTGTAGGGT	AGA		
15	1750	1770	1790			
	GGGGTGGGGACAGAAACA	ACTTTCTCCTACTCTTCCTT	CCTAATAGGCATCTCCAGG	CTG		
20	1810	1830	1850			
	CCTGGTCACTGCCCCTCT	CTCAGTGTCAATAGATGAAA	GTACATTGGGAGTCTGTAG	GAA		
25	1870	1890	1910			
	ACCCAACCTTCTTGTCAT	TGAAATTTGGCAAAGCTGAC	TTTGGGAAAGAGGGACCAG	AAC		
30	1930	1950	1970			
	TTCCCCTCCCTTCCCCTT	TTCCCAACCTGGACTTGTTT	TAAACTTGCCTGTTCAGAG	CAC		
35	1990	2010	2030			
	TCATTCCTTCCCACCCCC	AGTCCTGTCCTATCACTCTA	ATTCGGATTTGCCATAGCC	TTG		
	2050	2070	2090			
40	AGGTTATGTCCTTTTCCA	TTAAGTACATGTGCCAGGAA	ACAGCGAGAGAGAAAGTA	AAA		
	2110	2130	2150			
45	CGGCAGTAATGCTTCTCC	CGGCAGTAATGCTTCTCCTATTTCTCCAAAGCCTTGTGTGAACTAGCAAAGAGAAAA				
	2170	2190	2210			
50	TCAAATATATAACCAATA	GTGAAATGCCACAGGTTTG7	CCACTGTCAGGGTTGTCTA	CCT		

(

	2230	2250	2270			
	GTAGGATCAGGGTCTAAG	CACCTTGGTGCTTAGCTAGA	ATACCACCTAATCCTTCTGGC	•		
5		2210	2330			
	2290	2310				
	AGCCTGTCTTCAGAGAAC	CCACTAGAAGCAACTAGGAA	AAATCACTTGCCAAAATCCAA	3		
10	2350	2370	2390			
	GCAATTCCTGATGGAAAA	TGCAAAAGCAĆATATATGTT	TTAATATCTTTATGGGCTCTG	r		
15	2410	2430	2450			
	TCAAGGCAGTGCTGAGAG	GGAGGGGTTATAGCTTCAGG	AGGGAACCAGCTTCTGATAAA	ċ		
20	2470	2490	2510			
	ACAATCTGCTAGGAACTT	GGGAAAGGAATCAGAGAGCT	CCCTTCAGCGATTATTTAAA	Ť		
25	2530	2550	2570			
	TGTTAAAGAATACACAAT	TTGGGGTATTGGGATTTTT	CTCCTTTTCTCTGAGACATTCC	À		
30	2590	2610	2630			
	CCATTTTAATTTTTGTAACTGCTTATTTTTTGTGAAAAGGGTTATTTTTACTTAGCTTAGC					
05	2650	2670	2690			
35	TATGTCAGCCAATCCGAT	TGCCTTAGGTGAAAGAAAC	CACCGAAATCCCTCAGGTCCCT	`T		
	2710	2730	2750			
40	GGTCAGGAGCCTCTCAAG	ATTTTTTTTGTCAGAGGCT	CCAAATAGAAAATAAGAAAAGO	T		
	2770	2790	2810			
45	TTTCTTCATTCATGGCTA	GAGCTAGATTTAACTCAGT	TTCTAGGCACCTCAGACCAAT	:Α		
	2830	2850	2870			
50	TCAACTACCATTCTATTC	CATGTTTGCACCTGTGCAT	TTTCTGTTTGCCCCCATTCAC	T		

	2890			
	_	2910	2930	
	TGTCAGGAAACCTTGGC	CTCTGCTAAGGTGTATTTGG	PCCTTGAGAAGTGGGAGCA	CCC
5	2050	•	•	
	2950	2970	2990	
	ACAGGGACACTATCACTO	CATGCTGGTGGCATTGTTTA	CAGCTAGAAAGCTGCACTG	GTG
10	3010			
		3030	3050	
	TAATGCCCCTTGGGAAAT	GGGGCTGTGAGGAGGAGGA	TATAACTTAGGCCTAGCC	rcti
15	3070	2000		
		3090	3110	
	TIANCAGCCTCTGAAATT	TATCTTTCTTCTATGGGG	CTATAAATGTATCTTATA	Aata
20	3130	3150	3170	
	AAAGGAAGGACAGGAGGA	AGACAGGCAAATGTACTTCT		Satg
	3190	3210	•	
25	•		3230	
	GARICICII I IGGGCTAA	GAGAAAGGTTTTATTCTA'FA	TTGCTTACCTGATCTCATC	ATT
30	3250	3270	3290	
30	GGCCTAAGAGGCTTTCTC	CAGGAGGATTAGCTTGGAGT		CTT
	;			
35	3310	3330	3350	
	TCAGGGTTTTCTAACCCT	GACACGGACTGTGCATACTT	TCCCTCATCCATGCTGTGC	TGT
	3370	3390	3410	
40	GTTATTTAATTTTTCCTG	GCTAAGATCATGTCTGAATT		
		our montened to to AATT	ATGTATGAAAATTATTCTA	TGT
	3430	3450	·	
45	TTTTATAAAAAAATAAT	ATATCAGACATCGAAAAAA	Lan	

(d)

CC 666 664 CAC 6CA 666 CCA ACA 61C ACA 6CA 6CC C16 ACC A6A 6CA 11C C16 GAG C1C 1 . • AAS CIC TCT ACA AAS ASS ISS ACA SAS AAS ACA SCA SAS ACC AIG SSA CCC CCC ICA Met Gly Pro Pro Ser ł ECC CCT CCC TEC AGA TIE CAT ETC CCC TEE AAE EAE ETC CTE CTC ACA ECC TCA CTT Ala Pro Pro Cys Arg Leu His Val Pro Trp Lys 61u Val Leu Leu Thr Ala Ser Leu CIA ACC TIC TGG RAC CCA CCC ACC ACT GCC AAG CTC ACT ATT GAA TCC ACG CCA TTC Leu Thr Phe Trp Asn Pro Pro Ihr Thr Ala Lys Leu Thr Ile Glu Ser Thr Pro Phe · 570 AAT STC SCA SAS SSS SAS STT CTT CTA CTC SCC CAC AAC CTS CCC CAG AAT CST Asn Val Ala Siu Siy tys Siu Val Leu Leu Leu Ala His Asn Leu Pro Sin Asn Arg 10 /1 /2 /3 /9 /5 16 17 /8 /9 20 21 22 23 24 25 26 27 22 ATT 661 TAC AGC 166 TAC ARA 66C GAA AGA 616 GAT 66C AAC AGT CTA ATT 61A 66A lie bly lyr Ser Irp lyr Lys bly blu Arg Val Asp bly Ash Ser Leu lie Val bly 27 30 3/ 37 33 37 35 36 37 33 59 90 11/ 43 43 49 45 12 47 TAT STA ATA SSA ACT CAA CAA SCT ACC CCA SSS CCC SCA TAC AST SST CSA SAS ACA Tyr Val lie bly lhr bla bla Ala lhr Pro bly Pro Ala lyr Ser bly Arg blu lhr 42 41 50 51 53 53 57 55 56 57 58 58 60 67 63 69 65 66 ATA TAC CCC AAT BCA TCC CTB CTB ATC CAB AAC BTC ACC CAB AAT BAC ACA BBA TTC lle Tyr Pro Asn Ala Ser Leu Leu ile bin Asn Val Ihr bin Asn Asp Ihr biy Phe 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 62 63 79 85 • THE ACC CTA CAR SIC ATA AND TER BAT CTT STE ART, BAR BAR BER ACC BER CAS TIE Tyr Thr Leu Sin Val lie Lys Ser Asp Leu Val Asn Siu Siu Ala Thr Siy Sin Phe 26 E7 E2 E7 10 17 VL V3 94 95 96 97 98 98 00 101 103 103 104 CAT STA TAC CCS SAS CTS CCC AAS CCC TCC ATC TCC AGC AAC ACC TCC AAC CCC STS His. Val lyr Pro Siu Leu Pro Lys Pro Ser Ile Ser Ser Asn Asn Ser Asn Pro Val 165 166 167 166 167 160 111 112 115 116 117 118 119 120 121 122 123

566 SAC AGE SEE SAT SECT STEE SECT TIC ACT IST SAA CIT SAE SET CAE SAC ACT TAE  SIN 105-127 205 313 307 AT ATT TO 105 514 510 510 510 410 150 ASS TO 107 177  430 640 650 650 650 610 470 680  C16 166 166 618 ANT SET CAE AGE CIT CTC STE AGT CTC AGE CTG CAE CTG TO 107 179 179 179 179 179 179 179 179 179 17		SBO	590	400	610	.20
10   10   15   16   16   16   16   16   16   16	5	blu Aso Lva	ASO ALI VI AI	a Pho the fue the	Pen 61, 11,1 61, 4.	
CIG 166 66 618 ART 651 CA6 ARC CIC CCG 61C ARC CIC CA6 CIG CA6 CIG CIG CIG CIC ART  LEW Try Try Cid San Sity Sin Sty Lew Pro Val Ser Pro 619 Lew Sin Lew Sty Art 717 Art 717 Art 717 Art 718 A		630	640	50 650	670	680
66C ARC AGG ACC CIC ACT CIC ACT CIC AGC GIC AAA AGG ARC GAI GCA GGA ICG IAI GAA  617 ASA ATQ IAIT LEV IAT LEV LEV SET VAI LYS ATQ ASA ACQ AAA ACQ AAT ACQ ATT	10	Ley Trp 1rp	G BIA AAT BBI CA	AGC CTC CCG GTC	AGT CCC AGG CTG CA	1AA 331 813 8
15   16   16   16   16   17   18   17   18   17   18   18   18						-
150 760 770 780 770 800 770 800  161 6AA ATA CLG AGC CCA GCG AGT GCC AGC GCC AGT GCC CCA GCC AGT GCC CCA GCC AGT GCC CCC AGT GCC GCC GCC AGT GCC GCC GCC AGT GCC GCC AGT GCC GCC AGT GCC GCC GCC GCC GCC GCC AGT GCC AGT GCC AGT GCC GCC GCC GCC GCC GCC AGT GCC AGT GCC AGT GCC GCC GCC GCC GCC GCC GCC GCC GCC G	15	ara aza vid	ACC CIC ACT CI	A CIC AGC BIC AAA	AGG AAC GAI GCA GG	A ICS TAT BAA
161 688 ATA CAG AAC CCA GCG AGT BCC AAC CGC AGT GCC CTG AGT CCC CTG AAT GTC  Cy; Siu 11e Sin Asn Pre All Ser All Ash Arg Ser Asp Pre Vi) Int Leu Asn Vi)  810 820 830 840 850  CIC TAT GGC CCA GAT GGC CCC ACC ATT TCC CCC TCA AAG GCC AAT TAC CGI CCA GGG  Leu Tyr Gly Pre Asp Git Pre The Tile Ser Pre Ser Lys All Ash Tyr Lag Pre Gly 20 20 20 20 20 20 20 20 20 20 20 20 20 2		750	- 760			_
SIO   SZO   S3O   SAO   SZO	20	TET BAA ATA Cys Elu 11e	CAG AAC CCA GC	1 Car 61 . Ara Ara	AST SAC CCA SIC ACI	DID TAR DID
CIC IAI GGC CCA GAT GGC CCC ACC ATT ICC CCC ICA AAB GCC AAT IAC CGI CCA 666  Leu Tyr Gly Pro Asp Gly Pro Thr Ile Ser Pro Ser Lys Als Ash Tyr Gro Pro Gly  BAO 870 880 890 900 910  GAA AAI CIG AAC CIC ICC IGC CAC GCA GCC ICI AAC CCA CCI GCA CAG IAC ICI IGG  GIU ASh Leu Ash Leu Ser Cys Nijs Als Als Ser Ash Pro Pro Als Gin Iyr Ser Ire  20 920 930 940 950 960 970  TIT AIC AAT GGG ACG TIC CAG CAA ICC ACA GAA GAG CIC III AIC CCC AAC AIC ACI  Phe Ile Ash Gly Thr Phe Gin Gin Ser Ihr Gin Glu Leu Phe Ile Pro Ash Ile Ihr  22 237 270 271 272 273 277 277 277 277 277 277 277 277			810 82	0 630	_	
### 100   880   870   900   910    GRA ART CTG RAC CTC TCC 16C CAC 6CA 6CC 1CT ARC CCA CCT 6CA CAG TAC TCT 16G    STU ASA Lew Asa Lew Ser Cys His All All Ser Asa Pro Pro All 6th Tyr Ser Trp  920   930   910   950   760   970    TIT ATC RAT 6GG AC6 TTC CAG CAA CAA CAA GAG CTC TTT ATC CCC ARC ATC ACC    Phe 11e Asa 6th Tyr Phe 6th 6th Ser Thr 6th 6th 6th Pro Asa Tte Thr  262 237 300 311 307 317 317 317 317 317 317 317 317 317 31	25	Leu Irr Bly	CCA SAT SGC CC	C ACC ATT TCC CCC	TCA AAB GCC AAT TAG	
GAA AAT CIG AAC CIC IEC CAC BCA BCC ICI AAC CCA CCI BCA CAB IAC ICI IGG  GIU ASA Leu ASA Leu Ser Cre His Ala Ala Ser Asa Pro Pro Ala Bin Irr Ser Irp  920 930 940 950 950 960 970.  III AIC AAT BGB ACB IIC CAB CAA ICC ACA CAA BAB CIC III AIC CCC AAC AIC ACI Phe lie Asa Bir Ihr Phe Bin Bin Ser Ihr Bin Biu Leu Phe lie Pro Asa lie Ihr  282 237 200 270 270 270 270 270 270 270 270 27			870	880 . 890	900 ·	910
111 AIC AAT 6G6 AC6 TIC CA6 CAA ICC ACA CAA 6A6 CIC 111 AIC CCC AAC AIC ACI Phe lie Asn Gly Ihr Phe Gin Gin Ser Ihr Gin Giu Leu Phe ile Pro Asn 11e Ihr 200 229 200 21 21 21 21 21 21 21 21 21 21 21 21 21		GAA AAT CIB	AAC CIC ICC ISI Asn Lew Ser Cyr	CAC BCA BCC ICI	ا AAC CCA CCI 6CA CAG Asa Pro Pro Ala 61a محق تلت ملك ملك	TAC TCT TGG
Phe lie Ash Bly Thr Phe Bin Bin Ser Thr Bin Biu Leu Phe lie Pro Ash lie Thr  26 257 260 371 272 273 277 275 277 275 275 275 275 275 275 275	30	920	130 1	140 .4	750 760	· 970,
### 1000		Phe lle Asn	61v The Pho 61s	. Blm Ser The Bla (	Ein Ian Sha lia Sec	Aca Ila The
Val Ash Ash Ser 617 Ser 17r Nel Cys 61n Ala His Ash Ser Ala 1hr 617 Leu Ash  277 252 252 262 262 262 262 262 262 262 262	35					
40 1030 1040 1050 1060 1070 1080  A66-ACC ACA SIC ACG ATG ATC ACA SIC ICI SGA AGI SCI CCI SIC CIC ICA SCI SIG  Arg The The Val The Nel 119 The Val See Siz See Ala Pro Val Leu See Ala Val  372 377 378 378 38 38 38 38 38 38 38 38 38 38 38 38 38		Val Asn Asn	Ser bly Ser Tyr	Ret Cys Sto Ala 1	tic Aca See Als the	Elw Lan Bea
AGG- ACC ACA SIC ACG AIG AIC ACA SIC ICI SGA AGI SCI CCI SIC CIC ICA SCI SIG  Arg Thr Thr Val Thr Kel lie Thr Val Ser Sly Ser Ala Pro Val Leu Ser Ala Val  272 277 278 278 288 288 288 288 288 288	40					
Arg The The Val The Kel lie The Val See Bly See Ala Pro Val Leu See Ala Val  Arg The The Val The Kel lie The Val See Bly See Ala Pro Val Leu See Ala Val  Arg The The Val The Kel lie The Val See Bly See Ala Pro Val Leu See Ala Val  Arg The The Val The Wal See Ala Val  Arg The The Val The Wal See Ala Val  Arg The The Val The Val See Ala Val  Arg The Val The Val The Val See Ala Val  Arg The Val The Val The Val See Ala Val  Arg The Val The Val The Val See Ala Val  Arg The Val The Val The Val See Ala Val  Arg The Val The Val See Ala Val  Arg The Val The Val The Val  Arg The Val The Val The Val  Arg The Val The Val The Val  Arg The		-	-	•	•	1
45 1090 1100 . 1110 1120 1130 1110.  4		Are The The	الا الا الله الله الله الله الله الله ا	The Val See Sty S	161 6C1 CC1 61C C1C ier Ala Pro Val Leu 127 227 528 30 30	Ser Ala Val
SCC ACC GIC GGC AIC ACG AII GGA GIG CIG GCC AGG GIG GCI CIG AIA TAG CAG CCC	45	1090	1100 .	1110 11	120 1130	
· · · · · · · · · · · · · · · · · · ·		SCC ACC STC	SEC ATE ACE ATT	SEA STE CTE SCC	ATA 213 173 212 224	TAR CAR FFF

		1150			1160			1170				1180				1170				
	381	161	ATT		GAT	ATT	TCA	SGA	ASA	C16	6CA	ĠAI	166	ACC	AGA	נננ	164	ATT	CII	
5	1200	***	ret.	1210	430	***	1221	0		11	230		!	1			1856	0		
	: CIA	<b>b</b> L1	LLI	LLA	AIL	LLA	111	IAI	ttt	A15	BAA	CCA	CIA	AAA	AÇA	ASE	101	EC1	CIS	
10			•																	
	CTC	CIE	AAG	כככ	TAT	AIG	CIG	SAS	ATE	<b>BAC</b>	AAC	ICA	A16	AAA	All	TAR	A66	ARA	JIA	
15		1	320		1	1330	÷		- 1340	)		13	350			1360			1370	
	CCT	CAS	500	TEA	188	616	160	CAC	1CA	6A6	ACT	TCA	נכו	AAC	IAS	AGA	CAĢ	6CA	ARC	
20			1:	380	•		1390			140	•		10	10		1	1450			
	160	AAA	CCA	nnC	CTC	111	CBC	116	6CA	66A	TGA	166	161	CAT	IAG	TAI	110	ACA	AGA	
25	1430	)		14	40		1	1450			1460	١		. 10	79		. 1	1480		
	461	AGC	110	AGA	666	TAA	CII	RAC	A5A	614	1CA	SAT	<b>[</b> ]\$	ici	161	CAA	166	CAA	133	
30	1490		0 .		1500		1510		1530				13	1530		1540				
·	111	ACA	TAA	AAI	AAS	CGA	ICC	Ш	A61	6CA	CCC	A61	5AC	IGA	CAT	TAS	CAG	CAT	CIT	
				0	)		1560		1		570		1580		15		590	90		
35	3 TAR	CAC	AGC	CEI	616	110	AAG	161	ACA	615	610	CII	110	AGA	611	6Gn	nal	ACI	CCA	
	1600			161	0			650			1630			164			1	650		
40		6AA	ATS			AAS			BAT	CCA			AAA			ARC	CAA	111	AAA	
		1660			167			1	680		!	1690			170			ì	710	
45	ALE		AAA	BAA	CAC	ASS	AGA	TIC	CAB	101	AC1	İBA	611	ABC	ATA	ATA	CAG	AAG	1.	
			1720													176	-			
50	CCT	CTA	-		CII	114	CAA	AAA	AST	AAC	EIS			101	GAT	611		CAA	181	

55

~~

1770 1780 1770 1800 1810 1820 ATT TAT TIE TOT BET TOT BIT TOO TIE TIC CAR TIT EAC AAA ACC CAC TET TOT TET 1830 1810 1850 1840 1270 1880 ATT GIA TIG CCC AGG 666 AGC TAT CAC IGT ACT IGT AGA GIG GIG CIG CIT TAA STE 1890 1700 1710 1920 1130 1940 CAT ANA TER CAN ATA ARA SEC ART TAS CIC TAT ARE TAR ARA ARA ARA ARA 1950 1960 ALA CAR ARE ARE ARE RAR CAR

A schematic relationship of the transmembrane CEA's, namely TM-1 (CEA-(c)), TM-2 (CEA-(e)), TM-3 (CEA-(f)) and TM-4 (CEA-(g)) is depicted in Fig. 1:

Assuming TM-1 is composed of five sections as depicted in Fig. 1, namely 10, 12, 14, 16 and 18, TM-2 differs from TM-1 in that the 100 amino acid (100 AA) section 14 is deleted and at splice point 20 between sections 12 and 16, surprisingly an extra amino acid, namely Asp occurs.

TM-3 is the same as TM-1 except that section 18 is truncated at splice point 22, i.e., a section of 70 amino acids is deleted and results in a new section made up of subsections 24 + 26. Surprisingly, however, six new amino acids (section 26) occur. Another example of formation of a novel amino acid sequence resulting from a deletion of nucleic acid sequence is for platelet derived growth factor-A.

TM-4 is the same as TM-2 up until the end of subsection 24.

5

10

15

20

30

Subsection 24 is contained in section 18 of TM-1 and TM-2, but is not depicted in Fig. 1 for TM-1 and TM-2.

Some CEA epitopes are unique. These are the epitopes which have been useful for distinguishing the various CEA-like antigens immunologically. Peptide epitopes are defined by the linear amino acid sequence of the antigen and/or features resulting from protein folding. The information required for protein folding is encoded in the primary amino acid sequence. Therefore, antigenic differences ultimately result from differences in the primary structure of the different CEA molecules. The differences residing in the CEA protein in the CEA species can thus be determined by determining the primary amino acid sequences. This can be most readily accomplished by cloning and sequencing each of the genes for CEA. To determine which gene products will be most useful for cancer diagnosis, unique probes can be selected for each gene and expression of each gene can be determined in different tumor types by nucleic acid hybridization techniques. The present invention provides a tool with which to identify potential genes coding for different members of the CEA family and to determine the theoretical primary amino acid sequences for them. Using the method of automated peptide synthesis, peptides can then be synthesized corresponding to unique sequences in these antigens. With these peptides, antibodies to these sequences can be produced which, in the intact CEA molecule, might not be recognized by the animal being immunized. Having accomplished this, advantage can then be taken of the differences in these antigens to generate specific immunoassays for the measurement of each antigen.

A wide variety of host/cloning vehicle combinations may be employed in cloning the double-stranded nucleic acid prepared in accordance with this invention. For example, useful cloning vehicles may consist of segments of chromosomal, non-chromosomal and synthetic DNA sequences, such as various known derivatives of SV40 and known bacterial plasmids, e.g., plasmids from E. coli including col E1, pCR1, pBR322, pMB89 and their derivatives, wider host range plasmids, e.g.,  $\overline{RP4}$ , and phage DNAs, e.g., the numerous derivatives of phage, e.g., NM989, and other DNA phages, e.g., M13 and  $\overline{Filamenteous}$  single-stranded DNA phages and vectors derived from combinations of plasmids and phage DNAs such as plasmids which have been modified to employ phage DNA or other expression control sequences or yeast plasmids such as the 2  $\mu$  plasmid or derivatives thereof. Useful hosts may include bacterial hosts such as strains of  $\overline{E}$ . coli, such as  $\overline{E}$ . coli HB 101,  $\overline{E}$ . coli X1776,  $\overline{E}$ . coli X2282,  $\overline{E}$ . coli MRCI and strains of

Pseudomonas, Bacillus subtilis, Bacillus stearothermophilus and other E. coli, bacilli, yeasts and other fungi, animal or plant hosts such as animal (including human) or plant cells in culture or other hosts. Of course, not all host/vector combinations may be equally efficient. The particular selection of host/cloning vehicle combination may be made by those of skill in the art after due consideration of the principles sat forth without departing from the scope of this invention.

Furthermore, within each specific cloning vehicle, various sites may be selected for insertion of the nucleic acid according to the present invention. These sites are usually designated by the restriction endonuclease which cuts them. For example, in pBR322 the Pstl site is located in the gene for beta-lactamase, between the nucleotide triplets that code for amino acids 181 and 182 of that protein. One of the two Hindll endonuclease recognition sites is between the triplets coding for amino acids 101 and 102 and one of the several Taq sites at the triplet coding for amino acid 45 of beta-lactamase in pBR322. In similar fashion, the EcoRI site and the PVUII site in this plasmid lie outside of any coding region, the EcoRI site being located between the genes coding for resistance to tetracycline and ampicillin, respectively. These sites are well recognized by those of skill in the art. It is, of course, to be understood that a cloning vehicle useful in this invention need not have a restriction endonuclease site for insertion of the chosen DNA fragment. Instead, the vehicle could be cut and joined to the fragment by alternative means.

The vector or cloning vehicle and in particular the site chosen therein for attachment of a selected nucleic acid fragment to form a recombinant nucleic acid molecule is determined by a variety of factors, e.g., the number of sites susceptible to a particular restriction enzyme, the size of the protein to be expressed, the susceptibility of the desired protein to proteolytic degradation by host cell enzymes, the contamination of the protein to be expressed by host cell proteins difficult to remove during purification, the expression characteristics, such as the location of start and stop codons relative to the vector sequences, and other factors recognized by those of skill in the art. The choice of a vector and an insertion site for a particular gene is determined by a balance of these factors, not all sections being equally effective for a given case.

Methods of inserting nucleic acid sequences into cloning vehicles to form recombinant nucleic acid molecules include, for example, dA-dT tailing, direct ligation, synthetic linkers, exonuclease and polymerase-linked repair reactions followed by ligation, or extension of the nucleic acid strand with an appropriate polymerase and an appropriate single-stranded template followed by ligation.

It should also be understood that the nucleotide sequences or nucleic acid fragments inserted at the selected site of the cloning vehicle may include nucleotides which are not part of the actual structural gene for the desired polypeptide or mature protein or may include only a fragment of the complete structural gene for the desired protein or mature protein.

30

The cloning vehicle or vector containing the foreign gene is employed to transform an appropriate host so as to permit that host to replicate the foreign gene and to express the protein coded by the foreign gene or portion thereof. The selection of an appropriate host is also controlled by a number of factors recognized by the art. These include, for example, the compatibility with the chosen vector, the toxicity of proteins encoded by the hybrid plasmid, the ease of recovery of the desired protein, the expression characteristics, biosafety and costs. A balance of these factors must be struck with the understanding that not all hosts may be equally effective for expression of a particular recombinant DNA molecule.

The level of production of a protein is governed by two major factors: the number of copies of its gene within the cell and the efficiency with which those gene copies are transcribed and translated. Efficiency of transcription and translation (which together comprise expression) is in turn dependent upon nucleotide sequences, normally situated ahead of the desired coding sequence. These nucleotide sequences or expression control sequences define inter alia, the location at which RNA polymerase interacts to initiate transcription (the promoter sequence) and at which ribosomes bind and interact with the mRNA (the product of transcription) to initiate translation. Not all such expression control sequences function with equal efficiency. It is thus of advantage to separate the specific coding sequences for the desired protein from their adjacent nucleotide sequences and fuse them instead to other known expression control sequences so as to favor higher levels of expression. This having been achieved, the newly engineered nucleic acid, e.g., DNA, fragment may be inserted into a multicopy plasmid or a bacteriophage derivative in order to increase the number of gene copies within the cell and thereby further improve the yield of expressed protein.

Several expression control sequences may be employed as described above. These include the operator, promoter and ribosome binding and interaction sequences (including sequences such as the Shine-Dalgarno sequences) of the lactose operon of  $E.\ coli\ ("the lac system")$ , the corresponding sequences of the tryptophan synthetase system of  $E.\ coli\ ("the trp\ system")$ , the major operator and promoter regions of phage  $\lambda\ (O_LP_L\ and\ O_RP'_R)$ , the  $Control\ region$  of  $Control\ regio$ 

their viruses. Therefore, to improve the production of a particular polypeptide in an appropriate host, the gene coding for that polypeptide may be selected and removed from a recombinant nucleic acid molecule containing it and reinserted into a recombinant nucleic acid molecule closer or in a more appropriate relationship to its former expression control sequence or under the control of one of the above described expression control sequences. Such methods are known in the art.

As used herein "relationship" may encompass many factors, e.g., the distance separating the expression enhancing and promoting regions of the recombinant nucleic acid molecule and the inserted nucleic acid sequence, the transcription and translation characteristics of the inserted nucleic acid sequence or other sequences in the vector itself, the particular nucleotide sequence of the inserted nucleic acid sequence and other sequences of the vector and the particular characteristics of the expression enhancing and promoting regions of the vector.

Further increases in the cellular yield of the desired products depend upon an increase in the number of genes that can be utilized in the cell. This is achieved, for illustration purposes, by insertion of recombinant nucleic acid molecules engineered into the temperate bacteriophage λ (NM989), most simply by digestion of the plasmid with a restriction enzyme, to give a linear molecule which is then mixed with a restricted phage λ cloning vehicle (e.g., of the type described by N. E. Murray et al, "Lambdoid Phages That Simplify the Recovery of In Vitro Recombinants", Molec. Gen. Genet., 150, pp. 53-61 (1977) and N. E. Murray et al, "Molecular Cloning of the DNA Ligase Gene From Bacteriophage T4", J. Mol. Biol., 132, pp. 493-505 (1979)) and the recombinant DNA molecule recircularized by incubation with DNA ligase. The desired recombinant phage is then selected as before and used to lysogenize a host strain of E. coli.

Particularly useful  $\lambda$  cloning vehicles contain a temperature-sensitive mutation in the repression gene cl and suppressible mutations in gene S, the product of which is necessary for lysis of the host cell, and gene E, the product of which is major capsid protein of the virus. With this system, the lysogenic cells are grown at 32 °C and then heated to 45 °C to induce excision of the prophage. Prolonged growth at 37 °C leads to high levels of production of the protein, which is retained within the cells, since these are not lysed by phage gene products in the normal way, and since the phage gene insert is not encapsulated it remains available for further transcription. Artificial lysis of the cells then releases the desired product in high yield.

In addition, it should be understood that the yield of polypeptides prepared in accordance with this invention may also be improved by substituting different codons for some or all of the codons of the present DNA sequences, these substituted codons coding for amino acids identical to those coded for by the codons replaced.

Finally, the activity of the polypeptides produced by the recombinant nucleic acid molecules of this invention may be improved by fragmenting, modifying or derivatizing the nucleic acid sequences or polypeptides of this invention by well-known means, without departing from the scope of this invention.

The polypeptides of the present invention include the following:

- (1) the polypeptides expressed by the above described cells,
- (2) polypeptides prepared by synthetic means,

30

35

40

(3) fragments of polypeptides (1) or (2) above, such fragments produced by synthesis of amino acids or by digestion or cleavage.

Regarding the synthetic peptides according to the invention, chemical synthesis of peptides is described in the following publications: S.B.H. Kent, Biomedical Polymers, eds. Goldberg, E.P. and Nakajima, A. (Academic Press, New York), 213-242, (1980); A.R. Mitchell, S.B.H. Kent, M. Engelhard and R.B. Merrifield, J. Org. Chem., 43, 2845-2852, (1978); J.P. Tam, T.-W. Wong, M. Riemen, F.-S. Tjoeng and R.B. Merrifield, Tet. Letters, 4033-4036, (1979); S. Mojsov, A.R. Mitchell and R.B. Merrifield, J. Org. Chem., 45, 555-560, (1980); J.P. Tam, R.D. DiMarchi and R.B. Merrifield, Tet. Letters, 2851-2854, (1981); and S.B.H. Kent, M. Riemen, M. Le Doux and R.B. Merrifield, Proceedings of the IV International Symposium on Methods of Protein Sequence Analysis, (Brookhaven Press, Brookhaven, NY), in press, 1981.

In the Merrifield solid phase procedure, the appropriate sequence of L-amino acids is built up from the carboxyl terminal amino acid to the amino terminal amino acid. Starting with the appropriate carboxyl terminal amino acid attached to a polystyrene (or other appropriate) resin via chemical linkage to a chloromethyl group, benzhydrylamine group, or other reactive group of the resin, amino acids are added one by one using the following procedure. The peptide-resin is:

- (a) washed with methylene chloride;
- (b) neutralized by making for 10 minutes at room temperature with 5% (v/v) diisopropylethylamine (or other hindered base) in methylene chloride;
- (c) washed with methylene chloride;
- (d) an amount of amino acid equal to six times the molar amount of the growing peptide chain is activated by combining it with one-half as many moles of a carbodiimide (e.g., dicyclohexylcarbodiimide,

or diisopropylcarbodiimide) for ten minutes at 0 °C, to form the symmetric anhydride of the amino acid. The amino acid used should be provided originally as the N-alpha-tert.-butyloxycarbonyl derivative, with side chains protected with benzyl esters (e.g., aspartic or glutamic acids), benzyl ethers (e.g., serine, threonine, cysteine or tyrosine), benzyloxycarbonyl groups (e.g., lysine) or other protecting groups commonly used in peptide synthesis;

- (e) the activated amino acid is reacted with the peptide-resin for two hours at room temperature, resulting in addition of the new amino acid to the end of the growing peptide chain;
- (f) the peptide-resin is washed with methylene chloride;

5

10

30

35

50

55

- (g) the N-alpha-(tert.-butyloxycarbonyl) group is removed from the most recently added amino acid by reacting with 30 to 65%, preferably 50% (v/v) trifluoroacetic acid in methylene chloride for 10 to 30 minutes at room temperature;
- (h) the peptide-resin is washed with methylene chloride;
- (i) steps (a) through (h) are repeated until the required peptide sequence has been constructed.

The peptide is then removed from the resin and simultaneously the side-chain protecting groups are removed, by reaction with anhydrous hydrofluoric acid containing 10% v/v of anisole or other suitable (aromatic) scavenger. Subsequently, the peptide can be purified by gel filtration, ion exchange, high pressure liquid chromatography, or other suitable means.

In some cases, chemical synthesis can be carried out without the solid phase resin, in which case the synthetic reactions are performed entirely in solution. The reactions are similar and well known in the art, and the final product is essentially identical.

Digestion of the polypeptide can be accomplished by using proteolytic enzymes, especially those enzymes whose substrate specificity results in cleavage of the polypeptide at sites immediately adjacent to the desired sequence of amino acids.

Cleavage of the polypeptide can be accomplished by chemical means. Particular bonds between amino acids can be cleaved by reaction with specific reagents. Examples include the following: bonds involving methionine are cleaved by cyanogen bromide; asparaginyl-glycine bonds are cleaved by hydroxylamine.

The present invention has the following advantages:

- (1) The nucleic acids coding for TM-1, TM-2 and TM-3 can be used as probes to isolate other members of the CEA gene family.
- (2) The nucleic acids coding for TM-1, TM-2 and TM-3 can be used to derive oligonucleotide probes to determine the expression of TM-1, TM-2, TM-3 and other CEA genes in various tumor types.
- (3) TM-1, TM-2, TM-3 and TM-4 nucleotide sequences can be used to predict the primary amino acid sequence of the protein for production of synthetic peptides.
- (4) Synthetic peptides derived from the above sequences can be used to produce sequence-specific antibodies.
- (5) Immunoassays for each member of the CEA antigen family can be produced with these sequencespecific antibodies and synthetic peptides.
- (6) The aforementioned immunoassays can be used as diagnostics for different types of cancer if it is determined that different members of the CEA family are clinically useful markers for different types of cancer.

Peptides according to the present invention can be labelled by conventional means using radioactive moieties (e.g., <sup>125</sup>I), enzymes, dyed and fluorescent compounds, just to name a few.

Several possible configurations for immunoassays according to the present invention can be used. The readout systems capable of being employed in these assays are numerous and non-limiting examples of such systems include fluorescent and colorimetric enzyme systems, radioisotopic labelling and detection and chemiluminescent systems. Two examples of immunoassay methods are as follows:

- (1) An enzyme linked immunoassay (ELISA) using an antibody preparation according to the present invention (including Fab or F(ab)' fragments derived therefrom) to a solid phase (such as a microtiter plate or latex beads) is attached a purified antibody of a specificity other than that which is conjugated to the enzyme. This solid phase antibody is contacted with the sample containing CEA antigen family members. After washing, the solid phase antibody-antigen complex is contacted with the conjugated antipeptide antibody (or conjugated fragment). After washing away unbound conjugate, color or fluorescence is developed by adding a chromogenic or fluorogenic substrate for the enzyme. The amount of color or fluorescence developed is proportional to the amount of antigen in the sample.
- (2) A competitive fluorometric immunoassay using fluorescently labelled peptide or synthetic peptides of the sequences for TM-2, TM-3 and TM-4. In this example, the purified peptide expressed by cells or synthetic peptides thereof are fluorescently labelled. To a solid phase is attached a purified antibody. This solid phase is then contacted with sample containing CEA antigen family members to which has

been added fluorescent peptide probe. After binding, excess probe is washed away the amount of bound probe is quantitated. The amount of bound fluorescent probe will be inversely proportional to the amount of antigen in the sample.

In the nucleic acid hybridization method according to the present invention, the nucleic acid probe is conjugated with a label, for example, an enzyme, a fluorophore, a radioisotope, a chemiluminescent compound, etc. In the most general case, the probe would be contacted with the sample and the presence of any hybridizable nucleic acid sequence would be detected by developing in the presence of a chromogenic enzyme substrate, detection of the fluorophore by epifluorescence, by autoradiography of the radioisotopically labelled probe or by chemiluminescence. The detection of hybridizable RNA sequences can be accomplished by (1) a dot blot methodology or (2) an in situ hybridization methodology. Methods for these last two techniques are described by D. Gillespie and J. Bresser, "mRNA Immobilization in Nal: Quick Blots", Biotechniques, 184-192, November/December 1983 and J. Lawrence and R. Singer, "Intracellular Localization of Messenger RNAs for Cytosketal Proteins", Cell, 45, 407-415, May 9, 1986, respectively. The readout systems can be the same as described above, e.g., enzyme labelling, radiolabelling, etc.

As stated above, the invention also relates to the use in medicine of the aforementioned complex of the invention.

The invention further provides a pharmaceutical composition containing as an active ingredient a complex of the invention in the form of a sterile and/or physiologically isotonic aqueous solution.

For parenteral administration, solutions and emulsions containing as an active ingredient the complex of the invention should be sterile and, if appropriate, blood-isotonic.

It is envisaged that the active complex will be administered perorally, parenterally (for example, intramuscularly, intraperitoneally, or intravenously), rectally or locally.

### Example 1: Preparation of cDNA in pcE22 which codes for TM2-CEA [CEA-(e)]

#### Example 1a: RNA Preparation

15

25

Messenger RNA was prepared by the proteinase K extraction method of J. Favolaro, R. Treisman and R. Kamen, Methods in Enzymology, 65, 718, Academic Press, Inc. (1980), followed by oligo dT cellulose chromatography to yield poly A+ RNA (3'-polyadenylated eukaryotic RNA containing most mRNA sequences that can be translated into polypeptides). To obtain approximately 100 µg of poly A+ RNA, approximately 3 x 108 cells of transfectant 23.411 (ATCC No. CRL 9731, deposited with the ATCC on June 1, 1988), that expresses TM-1, TM-2, TM-3 and TM-4, Kamarck et al, Proc. Natl. Acad. Sci., USA, 84, 5350-5354, August 1987, were harvested from roller bottles after late logarithmic growth. Cells were lysed by homogenization in an ice-cold solution of 140 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 8.0, 0.5% NP40®, 4 mM dithiothreitol and 20 units of placental ribonuclease inhibitor/ml. Sodium deoxycholate was then added to 0.2%. Cytoplasm and nuclei were separated by centrifugation of the homogenate at 12,000xg for 20 minutes. The cytoplasmic fraction was mixed with an equal volume of 0.2 M Tris-HCl, pH 7.8, 25 mM EDTA, 0.3 M NaCl, 2% sodium dodecyl sulfate and 400 μg/ml of proteinase K, incubated for 1 hour at 37 °C, then extracted once with an equal volume of phenol/cholorform (1:1/v:v) solution. Nucleic acids were obtained by ethanol precipitation of the separated aqueous phase. Total RNA was enriched by passage in 0.5 M NaCl, 10 mM Tris-HCl, pH 7.8, 0.1% sarcosyl® through an oligo dT(12-18) cellulose column. After washing, bound RNA was eluted in the same solution without sodium chloride.

### Example 1b: Reverse Transcription of mRNA

Ten micrograms of poly A+ RNA were primed for reverse transcription with oligo dT(12-18) and pdN<sub>6</sub> primers. One hundred microliter reaction was performed for 4 hours at 42 °C with 200 units AMV reverse transcriptase (Life Science, Inc. St. Petersburg, Florida, U.S.A.). The RNA component of the cDNA/mRNA hybrids was replaced with the second complementary strand by treatment with RNase H, E. coli DNA polymerase I and E. coli DNA ligase at 12 °C and 22 °C for 1.5 hours each. Molecular ends were polished by treatment with T4 DNA polymerase. cDNA was phenol/chloroform extracted and purified over a "SEPHADEX® G-50" spun column prepared in 10 mM Tris-HCl, pH 7.8, 1 mM EDTA (TE).

### Example 1c: Cloning of pcE22 (plasmid cDNA E22)

Synthetic DNA linkers 5' pCCCGGG 3' 3' GGGCCCTTAA 5'

were attached to the ends of cDNA by blunt end ligation with excess T4 DNA ligase. Excess linkers were removed by chromatography through "SEPHADEX® G-50" (medium) in TE, and by fractionation on 0.8% low melting agarose gel. Based on Northern blot analysis of poly A + RNA of the 23.411 cell line, the size of the CEA-related mRNA was estimated at 3.6 kb. Therefore, cDNA fragments between 2 and 4 kb were recovered from gel slices and fragments were ethanol precipitated. After resuspension of cDNA in TE, EcoRI-cleaved lambda gt10 arms were added to cDNA at an estimated molar ratio of 1:1. Ligation proceeded at 7 °C for 2 days in the presence of T4 DNA ligase. Aliquots of the ligation reaction were added to commercially-obtained packaging mix (Stratagene, San Diego, California, U.S.A.). Five million phage particles were obtained ofter in vitro packaging and infection of E. coli host NM514.

### Example 1d: Screening of Recombinant Library

Five hundred thousand packaged lambda particles were plated on lawns of E. coli NM514 and replicate patterns were lifted onto nitrocellulose sheets as described by W.D. Benton and R.W. Davis, Science 196, 180-182, (1977). Positive phage were selected by hybridization with <sup>32</sup>P-labeled LV7 cDNA insert probe that contained a domain repeated amoung various CEA family members. By multiple rounds of screening. Phage from individual plaques were amplified and titered, and these were used to prepare small quantities of recombinant phage DNA.

### Example 1e: DNA Manipulation

10

30

50

Phage DNA was prepared according to T. Maniatis, E. Fritsch and J. Sambrook, Molecular Cloning, A Laboratory Manual, Cold Spring Habor, (1982). DNA segments were isolated from low melting agarose gels and inserted for subcloning into Bluescript plasmid vectors (Stratagene, San Diego, California, U.S.A.). DNA sequencing was performed by the dideoxy termination method of F. Sanger, S. Nicklen and A. Coulson, Proc. Natl. Acad. Sci., U.S.A., 74, 5463-5467, (1977). The nucleic acid and translated sequence for cDNA in pcE22 is given hereinabove (TM-2 (CEA-(e)).

### Example 2: Preparation of cDNA in pcHT-6 which Partically Codes for TM3-CEA [CEA-(f)]

#### Example 2a: RNA Preparation

Messenger RNA was prepared by the proteinase K extraction method of J. Favolaro, R. Treisman and R. Kamen, Methods in Enzymology, 65 718, Academic Press, Inc. (1980), followed by oligo dT cellulose chromatography to yield poly A+ RNA (3'-polyadenylated eukaryotic RNA containing most mRNA sequences that can be translated into polypeptides). To obtain approximately 100 ug of poly A+ RNA, approximately 3 x 10<sup>8</sup> cells of HT-29 tumor cells (ATCC HTB38) were harvested form roller bottles after late logarithmic growth. Cells were lysed by homogenization in an ice-cold solution of 140 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 8.0, 0.5% NP40®, 4 mM dithiothreitol and 20 units of placental ribonuclease inhibitor/ml. Sodium deoxycholate was then added to 0.2%. Cytoplasm and nuclei were separated by centrifugation of the homogenate at 12,000xg for 20 minutes. The cytoplasmic fraction was mixed with an equal volume of 0.2 M Tris-Hcl, pH 7.8, 25 mM EDTA, 0.3 M NaCl, 2% sodium dodecyl sulfate and 400 μg/ml of proteinase K, incubated for 1 hour at 37°C, then extracted once with an equal volume of phenol/cholorform (1:1/v:v) solution. Nucleic acids were obtained by ethanol precipitation of the separated aqueous phase. Total RNA was enriched by passage in 0.5 M NaCl, 10 mM Tris-HCl, pH 7.8, 0.1% sarcosyl® through an oligo dT(12-18) cellulose column. After washing, bound RNA was eluted in the same solution without sodium chloride.

### Example 2b: Reverse Transcription of mRNA

Ten micrograms of HT-29 poly A+ RNA were primed for reverse transcription with oligo dT(12-18) and pdN₅ primers. One hundred microliter reaction was performed for 4 hours at 42 °C with 200 units AMV reverse transcriptase (Life Science, Inc. St. Petersburg, Florida, U.S.A.). The RNA component of the cDNA/mRNA hybrids was replaced with the second complementary strand by treatment with RNase H, E. coli DNA polymerase I and E. coli DNA ligase at 12 °C and 22 °C for 1.5 hours each. Molecular ends were polished by treatment with T4 DNA polymerase. cDNA was phenol/chloroform extracted and purified over a "SEPHADEX® G-50" spun column prepared in 10 mM Tris-HCl, pH 7.8, 1 mM EDTA (TE).

### Example 2c: Cloning of pcHT-6 (plasmid cDNA HT-6)

Synthetic DNA linkers 5' pCCCGGG 3' 3' GGGCCCTTAA 5'

were attached to the ends of cDNA by blunt end ligation with excess T4 DNA ligase. Excess linkers were removed by chromatography through "SEPHADEX® G-50" (medium) in TE, and by fractionation on 0.8% low melting agarose gel. Based on Northern blot analysis of poly A + RNA of the HT-29 cell line, the size of the CEA-related mRNA was estimated at 2.2 kb. Therefore, cDNA fragments between 2 and 3 kb were recovered from gel slices and fragments were ethanol precipitated. After resuspension of cDNA in TE, EcoRI-cleaved lambda gt10 arms were added to cDNA at an estimated molar ratio of 1:1. Ligation proceeded at 7°C for 2 days in the presence of T4 DNA ligase. Aliquots of the ligation reaction were added to commercially-obtained packaging mix (Stratagene, San Diego, California, U.S.A.). Five million phage particles were obtained ofter in vitro packaging and infection of E. coli host NM514.

### 5 Example 2d: Screening of Recombinant Library

Five hundred thousand packaged lambda particles were plated on lawns of <u>E. coli NM514</u> and replicate patterns were lifted onto nitrocellulose sheets as described by W.D. Benton and R.W. Davis, <u>Science</u>, 196, 180-182, (1977). Positive phage were selected by hybridization with <sup>32</sup>P-labeled LV7 cDNA insert probe that contained a domain repeated amoung various CEA family members. By multiple rounds of screening. Phage from individual plaques were amplified and titered, and these were used to prepare small quantities of recombinant phage DNA.

#### Example 2e: DNA Manipulation

25

Phage DNA was prepared according to T. Maniatis, E. Fritsch and J. Sambrook, Molecular Cloning, A Laboratory Manual, Cold Spring Habor, (1982). DNA segments were isolated from low melting agarose gels and inserted for subcloning into Bluescript plasmid vectors (Stratagene, San Diego, California, U.S.A.). DNA sequencing was performed by the dideoxy termination method of F. Sanger, S. Nicklen and A. Coulson, Proc. Natl. Acad. Sci., U.S.A., 74, 5463-5467, (1977). The nucleic acid and translated sequence for cDNA in HT-6 not complete at the 5' end of its coding region, but the nucleotide sequence and restriction map of the HT-6 insert indicates that it is related to nucleic acid sequences of cDNA clones coding for CEA-(c) and CEA-(e). The nucleotide sequence of HT-6 insert differs from these clones at only nucleotide position 1463 to 1515 and 1676 to 2429 of the CEA-(c) cDNA. It is inferred from this result that the pcHT-6 insert is a partial coding sequence for CEA-(f) and the presumed nucleic acid and translated sequence of CEA-(f) is given hereinabove (TM-3 (CEA-(f)).

### Example 3: Preparation of cDNA which codes for TM4-CEA [CEA-(g)]

#### Example 3a: RNA Preparation

Messenger RNA is prepared by the proteinase K extraction method of J. Favolaro, R. Treisman and R. Kamen, Methos in Enzymology, 65, 718, Academic Press, Inc. (1980), followed by oligo dT cellulose chromatography to yield poly A+ RNA (3'-polyadenylated eukaryotic RNA containing most mRNA sequences that can be translated into polypeptides). To obtain approximately 100 ug of poly A+ RNA, approximately 3 x 10<sup>8</sup> cells of transfectant 23.411 or tumor cell line HT-29 (ATCC HTB 38) are harvested from roller bottles after late logarithmic growth. Cells are lysed by homogenization in an ice-cold solution of 140 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 8.0, 0.5% NP40®, 4 mM dithiothreitol and 20 units of placental ribonuclease inhibitor/ml. Sodium deoxycholate was then added to 0.2%. Cytoplasm and nuclei are separated by centrifugation of the homogenate at 12,000xg for 20 minutes. The cytoplasmic fraction is mixed with an equal volume of 0.2 M Tris-Hcl, pH 7.8, 25 mM EDTA, 0.3 M NaCl, 2% sodium dodecyl sulfate and 400 μg/ml of proteinase K, incubated for 1 hour at 37°C, then extracted once with an equal volume of phenol/cholorform (1:1/v:v) solution. Nucleic acids are obtained by ethanol precipitation of the separated aqueous phase. Total RNA is enriched by passage in 0.5 M NaCl, 10 mM Tris-HCl, pH 7.8, 0.1% sarcosyl through an oligo dT(12-18) cellulose column. After washing, bound RNA is eluted in the same solution without sodium chloride.

#### Example 3b: Reverse Transcription of mRNA

Ten micrograms of 23.411 or HT 29 poly A+ RNA are primed for reverse transcription with oligo dT(12-18) and pdN<sub>6</sub> primers. One hundred microliter reaction was performed for 4 hours at 42°C with 200 units AMV reverse transcriptase (Life Science, Inc. St. Petersburg, Florida, U.S.A.). The RNA component of the cDNA/mRNA hybrids is replaced with the second complementary strand by treatment with RNase H, E. coli DNA polymerase I and E. coli DNA ligase at 12°C and 22°C for 1.5 hours each. Molecular ends are polished by treatment with T4 DNA polymerase. cDNA is phenol/chloroform extracted and purified over a "SEPHADEX® G-50" spun column prepared in 10 mM Tris-HCl, pH 7.8, 1 mM EDTA (TE).

#### Example 3c: Cloning of cDNA for CEA-(g)

10

Synthetic DNA linkers 5' pCCCGGG 3' 3' GGGCCCTTAA 5'

are attached to the ends of cDNA by blunt end ligation with excess T4 DNA ligase. Excess linkers are removed by chromatography through "SEPHADEX® G-50" (medium) in TE, and by fractionation on 0.8% low melting agarose gel. Based on Northern blot analysis of poly A + RNA of the 23.411 and HT-29 cell lines, the size of the CEA-related mRNA is estimated at 1.7 kb. Therefore, cDNA fragments between 1 and 2 kb are recovered from gel slices and fragments are ethanol precipitated. After resuspension of cDNA in TE, EcoRI-cleaved lambda gt10 arms are added to cDNA at an estimated molar ratio of 1:1. Ligation proceeds at 7°C for 2 days in the presence of T4 DNA ligase. Aliquots of the ligation reaction are added to commercially-obtained packaging mix (Stratagene, San Diego, California, U.S.A.). Phage particles are obtained after in vitro packaging and infection of E. coli host NM514.

#### Example 3d: Screening of Recombinant Library

Five hundred thousand to one million packaged lambda particles are plated on lawns of E. coli NM514 and replicate patterns are lifted onto nitrocellulose sheets as described by W.D. Benton and R.W. Davis, Science, 196, 180-182, (1977). Positive phage are selected by hybridization with <sup>32</sup>P-labeled LV7 cDNA insert probe that contained a domain repeated amoung various CEA family members. By this selection method, positive phage are obtained after multiple rounds of screening. Phage from individual plaques are amplified and titered, and these are used to prepare small quantities of recombinant phage DNA.

#### Example 3e: DNA Manipulation

Phage DNA is prepared according to T. Maniatis, E. Fritsch and J. Sambrook, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor, (1982). DNA segments are isolated from low melting agarose gels and inserted for subcloning into Bluescript plasmid vectors (Stratagene, San Diego, California, U.S.A.). DNA sequencing is performed by the dideoxy termination method of F. Sanger, S. Nicklen and A. Coulson, Proc. Natl. Acad. Sci., U.S.A., 74, 5463-5467, (1977). The nucleotide and translated sequence for a cDNA coding for CEA-(g) is given hereinabove (TM-4 (CEA-(g)).

#### Example 4: Screening of a KG-1 cDNA Library with <sup>32</sup>P-labelled CEA Probe, LV7 (CEA-(A))

A segment of cDNA coding for a portion of carcinoembryonic antigen [LV7 or CEA-(a)] was radiolabelled by random priming and used to detect homologous sequences on filter replicas of a commercial cDNA library prepared from KG-1 cells in bacteriophage vector λ gt11 (Clontech Laboratories, Inc., Palo Alto, CA., U.S.A.). Hybridizations were performed at 68 °C in 2xSSSPE (1xSSPE - 0.15 M NaCl, 0.01 M sodium phosphate and 1 mM EDTA, pH 7), 5x Denhardt's solution and 100 μg of denatured salmon sperm DNA per ml, and post-hybridization washes were in 0.2xSSC, 0.25% sodium dodecyl sulfate.

Positive phage were picked, rescreened to homogeneity, and amplified for production of DNA. cDNA inserts were excised from phage DNA with EcoRI endonuclease and subcloned into the EcoRI site of the plasmid vector pBluescript KS. DNA sequencing on double-stranded DNA was by the method of Sanger et al, supra. The sequences of two different inserts from the KG-1 cDNA library are shown below:

рс	K	G	C	EA	1	
----	---	---	---	----	---	--

	1	acagcacagctgacagccgtactcaggaagcttctggatcctaggcttatctccacagag	60
5	61	gagaacacacaagcagagaccatggggcccctctcagcccctccct	120
	121	atcacttggaaggggtcctgctcacagcatcacttttaaacttctggaatccgcccaca IleThrTrpLysGlyValLeuLeuThrAlaSerLeuLeuAsnPheTrpAsnProProThr	180
10	181	actgcccaagtcacgattgaagcccagccacccaaagtttctgaggggaaggatgttctt ThrAlaGlnValThrIleGluAlaGlnProProLysValSerGluGlyLysAspValLeu	240
	241	ctacttgtccacaatttgccccagaatcttgctggctacatttggtacaaagggcaaatg LeuLeuValHisAsnLeuProGlnAsnLeuAlaGlyTyrIleTrpTyrLysGlyGlnMet	300
15	301	acatacgtctaccattacattacatcatatgtagtagacgqtcaaagaattatatatggg ThrTyrValTyrHisTyrIleThrSerTyrValValAspGlyGlnArgIleIleTyrGly	360
	361	cctgcatacagtggaagagaaagagtatattccaatgcatccctgctgätccagaatgtc ProAlaTyrSerGlyArgGluArgValTyrSerAsnAlaSerLeuLeuIleGlnAsnVal	420
20	421	acgcaggaggatgcaggatcctacaccttacacatcataaagcgacgcgatgggactgga ThrGlnGluAspAlaGlySerTyrThrLeuHisIleIleLysArgArgAspGlyThrGly	480
	481	ggagtaactggacatttcaccttcaccttacacctggagactcccaagccctccatctcc GlyValThrGlyHisPheThrPhaThrLeuHisLeuGluThrProLysProSerIleSer	540
25	541	agcagcaacttaaatcccagggaggccatggaggctgtgatcttaacctgtgatcctgcg SerSerAsnLeuAsnProArgGluAlaMetGluAlaValIleLeuThrCysAspProAla	600
	601	actccagccgcaagctaccagtggtggatgaatggtcagagcctccctatgactcacagg ThrProAlaAlaSerTyrGlnTrpTrpMetAsnGlyGlnSerLeuProMetThrHisArg	660
30	661	ttgcagctgtccaaaaccaacaggaccctctttatattttggtgtcacaaagtatattgca LeuGlnLeuSerLysThrAsnArgThrLeuPheIlePheGlyValThrLysTyrIleAla	720
	721	ggaccctatgaatgtgaaatacggaacccagtgagtgccagccgcagtgacccagtcacc GlyProTyrGluCysGluIleArgAsnProValSerAlaSerArgSerAspProValThr	780
35	781	ctgaatctcctcccaaagctgtccaagccctacatcacaatcaacaacttaaaccccaga LeuAsnLeuLeuProLysLeuSerLysProTyrIleThrIleAsnAsnLeuAsnProArg	840
	841	gagaataaggatgtcttaaccttcacctgtgaacctaagagtgagaactacacctacatt GluAsnLysAspValLeuThrPheThrCysGluProLysSerGluAsnTyrThrTyrIle	900
40	901	tggtggctaaatggtcagagcctccctgtcagtcccagggtaaagcgacccattgaaaac TrpTrpLeuAsnGlyGlnSerLeuProValSerProArgValLysArgProIleGluAsn	960
	961	aggatecteattetacceaatgteaegagaaatgaaaeaggaeettateaatgtgaaata ArgIleLeuIleLeuProAsnValThrArgAsnGluThrGlyProTyrGlnCysGluIle	1020
45	1021	Cgggaccgatatggtggcatccgcagtgacccagtcaccctgaatgtcctctatggtcca	1080

	1081	gacctccccagcatttacccttcattcacctattaccgttcaggagaaaacctctacttt AspLeuProSerIleTyrProSerPheThrTyrTyrArgSerGlyGluAsnLeuTyrPhe	1140
5	1141	tcctgcttcggtgagtctaacccacgggcacaatattcttggacaattaatgggaagttt SerCysPheGlyGluSerAsnProArgAlaGlnTyrSerTrpThrIleAsnGlyLysPhe	1200
	1201	cagctatcaggacaaaagctctctatcccccaaataactacaaagcatagtgggctctat GlnLeuSerGlyGlnLysLeuSerIleProGlnIleThrThrLysHisSerGlyLeuTyr	1260
10	1261	gcttgctctgt&cgtaactcagccactggcaaggaaagctccaaatccatcacagtcaaa AlaCysSerValArgAsnSerAlaThrGlyLysGluSerSerLysSerIleThrValLys	1320
	1321	gtctctgactggatattaccctgaattctactagttcctccaattccattttctcccatg ValSerAspTrpIleLeuProEnd	1380
15	1 1 1 1 1 5 0 1 1 5 6 1 1 6 2 1 1 6 8 1 1 7 4 1	gaatcacgaagagcaagacccactctgttccagaagccctataatctggaggtggacaac tcgatgtaaatttcatgggaaaacccttgtacctgacatgtgagccactcagaactcacc aaaatgttcgacaccataacaacagctactcaaactgtaaaccaggataagaagttgatg acttcacactgtggacagtttttcaaagatgtcataacaagactccccatcatgacaagg ctccaccctctactgtctgctcatgcctgcctcttcacttggcaggataatgcagtcat tagaatttcacatgtagtagcttctgagggtaacaacagagtgtcagatatgtcatctca acctcaaacttttacgtaacatctcagggaaatgtggctctctccatcttgcatacaggg	1440 1500 1560 1620 1680 1740 1800
20	1801 1861 1921 1981	ctcccaatagaaatgaacacagagatattgcctgtgttgtttgcagagaagatggtttcta taaagagtaggaaagctgaaattatagtagagtctcctttaaatgcacattgtgtggatg gctctcaccatttcctaagagatacagtgtaaaaacgtgacagtaatactgattctagca gaataaacatgtaccacatttgcaaaaaa	1860 1920 1980 2010
25		pckgcea2:	
	1	gggtggatcctaggctcatctccataggggagaacacacatacagcagagaccatggga MetGly	59
30	60	cccctctcagcccctccctgcactcagcacatcacctggaaggggctcctgctcacagca ProLeuSerAlaProProCysThrGlnHisIleThrTrpLysGlyLeuLeuLeuThrAla	119
	120	tcacttttaaacttctggaacctgcccaccactgcccaagtaataattgaagcccagcca SerLeuLeuAsnPheTrp&snLeuProThrThrAlaGlnValIleIleGluAlaGlnPro	179
35	180	cccaaagtttctgaggggaaggatgttcttctacttgtccacaatttgccccagaatctt ProLysValSerGluGlyLysAspValLeuLeuLeuValHisAsnLeuProGlnAsnLeu	239
	240	actggctacatctggtacaaagggcaaatgacggacctctaccattacattacatcatat ThrGlyTyrIleTrpTyrLysGlyGlnMetThrAspLeuTyrHisTyrIleThrSerTyr	299
40	300	gtagtagacggtcaaattatatatgggcctgcctacagtggacgagaaacagtatattcc ValValAspGlyGlnIleIleTyrGlyProAlaTyrSerGlyArgGluThrValTyrSer	359
	360	aatgcatccctgctgatccagaatgtcacacagga <del>g</del> gatgcaggatcctacaccttacac AsnAlaSerLeuLeuIleGlnAsnValThrGlnGluAspAlaGlySerTyrThrLeuHis	419
45	420	atcataaagcgaggcgatgggactggaggagtaactggatatttcactgtcaccttatac IleIleLysArgGlyAspGlyThrGlyGlyValThrGlyTyrPheThrValThrLeuTyr	479
	480	tcggagactcccaagcgctccatctccagcagcaacttaaaccccagggaggtcatggag SerGluThrProLysArgSerIleSerSerSerAsnLeuAsnProArgGluValMetGlu	539

	540	gctgtgcgcttaatctgtgatcctgagactccggatgcaagctacctgtggttgctgaat AlaValArgLeuIleCysAspProGluThrProAspAlaSerTyrLeuTrpLeuLeuAsn	599
5	600	ggtcagaacctccctatgactcacaggttgcagctgtccaaaaccaacaggaccctctat GlyGlnAsnLeuProMetThrHisArgLeuGlnLeuSerLysThrAsnArgThrLeuTyr	659
	660	ctatttggtgtcacaaagtatattgcagggccctatgaatgtgaaatacggaggggagtg LeuPheGlyValThrLysTyrIleAlaGlyProTyrGluCysGluIleArgArgGlyVal	719
10	720	agtgccagccgcagtgacccagtcaccctgaatctcctcccgaagctgcccatgccttac SerAlaSerArgSerAspProValThrLeuAsnLeuLeuProLysLeuProMetProTyr	779
	780	atcaccatcaacaacttaaaccccaggqagaagaaggatgtgttagccttcacctgtgaa IleThrIleAsnAsnLeuAsnProArgGluLysLysAspValLeuAlaPheThrCysGlu	839
15	840	cctaagagtcggaactacacctacatttggtggctaaatggtcagagcctcccggtcagt ProLysSerArgAsnTyrThrTyrIleTrpTrpLeuAsnGlyGlnSerLeuProValSer	899
	900	ccgagggtaaagcgacccattgaaaacaggatactcattctacccagtgtcacgagaaat ProArgValLysArgProIleGluAsnArgIleLeuIleLeuProSerValThrArgAsn	959
2 <b>0</b>	960	gaaacaggaccctatcaatgtgaaatacgggaccgatatggtggcatccgcagtaaccca GluThrGlyProTyrGlnCysGluIleArgAspArgTyrGlyGlyIleArgSerAsnPro	1019
	1020	gtcaccctgaatgtcctctatggtccagacctccccagaatttacccttacttcacctat ValThrLeuAsnValLeuTyrGlyProAspLeuProArgIleTyrProTyrPheThrTyr	1079
25	1080	taccgttcaggagaaaacctcgacttgtcctgctttgcggactctaacccaccggcagag TyrArgSerGlyGluAsnLeuAspLeuSerCysPheAlaAspSerAsnProProAlaGlu	1139
	1140	tatttttggacaattaatgggaagtttcagctatcaggacaaaagctctttatcccccaa TyrPheTrpThrIleAsnGlyLysPheGlnLeuSerGlyGlnLysLeuPheIleProGln	1199
30	1 ,0	attactacaaatcatagcgggctctatgcttgctctgttcgtaactcagccactggcaag IleThrThrAsnHisSerGlybeuTyrAlaCysSerValArgAsnSerAlaThrGlyLys	1259
-	1260	gaaatotocaaatocatgatagtoaaagtototggtoootggcatggaaaccagacagag GluIleSerLysSerMetIleValLysValSerGlyProCysHisGlyAsnGlnThrGlu	1319
35	1320	tctcattaatggctgccacaatagagacactgagaaaaagaacaggttgataccttcatg SerHisEnd	1379
-•	1380 1440 1500 1560	aaattcaagacaaagaagaaaaaggctcaatgttattggactaaataatcaaaaggataa tgttttcataaatttttattggaaaatgtgctgattcttggaatgttttattctccagatt tatgaacttttttttcttcagcaattggtaaagtatacttttgtaaacaaaaattgaaaca tttgcttttgctctctatctgagtgcccccc 1591	1439 1499 1559

It will be appreciated that the instant specification and claims are set forth by way of illustration and not limitation and that various modifications and changes may be made without departing from the scope of the present invention.

## Claims

1. A nucleic acid comprising a base sequence which codes for a peptide sequence, characterized in that the group nucleic acid is a DNA selected from the following group of five sequences:

50

40

	490	510	530
5	CAGTTCCATGTATACCCGG GlnPheHisValTyrProG	AGCTGCCCAAGCCCTCCATCT luLeuProLysProSerIleS	CCAGCAACAACTCCAACCCT SerSerAsnAsnSerAsnPro
	550	570	590
10	GTGGAGGACAAGGATGCTG ValGluAspLysAspAlaV	TGGCCTTCACCTGTGAACCTC	GAGACTCAGGACACAACCTAC GluThrGlnAspThrThrTyr
	610	630	650
15	CTGTGGTGGATAAACAATC LeuTrpTrpIleAsnAsnC	AGAGCCTCCCGGTCAGTCCC SlnSerLeuProValSerPro	AGGCTGCAGCTGTCCAATGGC ArgLeuGlnLeuSerAsnGly
20	670	690	710
	AACAGGACCCTCACTCTAC AsnArgThrLeuThrLeut	CTCAGTGTCACAAGGAATGAC LeuSerValThrArgAsnAsp	ACAGGACCCTATGAGTGTGAA ThrGlyProTyrGluCysGlu
25	730	750	770
	ATACAGAACCCAGTGAGT IleGlnAsnProValSer	GCGAACCGCAGTGACCCAGTC AlaAsnArgSerAspProVal	CACCTTGAATGTCACCTATGGC ThrLeuAsnValThrTyrGly
30	790	810	830
35	CCGGACACCCCCACCATT ProAspThr?roThrIle	TCCCCTTCAGACACCTATTA( SerProSerAspThrTyrTy	CCGTCCAGGGGCAAACCTCAGC rArgProGlyAlaAsnLeuSer
	850	870	890
40	CTCTCCTGCTATGCAGCC LeuSerCysTyrAlaAla	TCTAACCCACCTGCACAGTA SerAsnProProAlaGlnTy	CTCCTGGCTTATCAATGGAACA rSerTrpLeuIleAsnGlyThr
	910	930	950
45	TTCCAGCAAAGCACACA/ PheGlnGlnSerThrGlr	AGAGCTCTTTATCCCTAACAT AGluLeuPheIleProAsnIl	CACTGTGAATAATAGTGGATCC eThrValAsnAsnSerGlySer
	970	990	1010
50	TATACCTGCCACGCCAA TyrThrCysHisAlaAs	TAACTCAGTCACTGGCTGCAA nAsnSerValThrGlyCysAs	.CAGGACCACAGTCAAGACGATC

(

	1030	1050	1070
5			CTCACCTGGGGCCATTGCTGGC userProGlyAlaIleAlaGly
	1090	1110	1130
10			AGTAGCCCTGGCATGTTTTCTG aValAlaLeuAlaCysPheLeu
	1150	1170	1190
15			TCTCACAGAGCACAAACCCTCA pLeuThrGluHisLysProSer
	1210	1230	1250
20			TAACAAGATGAATGAAGTTACT OASnLysMetAsnGluValThr
	1270	1290	1310
25			ACCAACTTCAGCCTCCCATCC .nProThrSerAlaSerProSer
30	1330	1350	1370
		ATTTATTCAGAAGTAAAAA IleTyrSerGluValLysLy	AGCAGTAATGAAACCTGTCCTGC ysGln
35	1390	1410	1430
	TCACTGCAGTGCTGATGT	`ATTTCAAGTCTCTCACCCT	CATCACTAGGAGATTCCTTTCCC
40	1450	1470	1490
	CTGTAGGGTAGAGGGGT	GGGACAGAAACAACTTTCT	CCTACTCTTCCTAATAGGC
45	1510	1530	1550
	ATCTCCAGGCTGCCTGG	rcactgcccctctctcagifg	TCAATAGATGAAAGTACATTGGG
	1570	1590	1610
50	AGTCTGTAGGAAACCCA	ACCTTCTTGTCATTGAAATT	TGGCAAAGCTGACTTTGGGAAAG

	1630	1650°	1670
5	AGGGACCAGAACTTCCCC	TCCCTTCCCCTTTTCCCAA	CCTGGACTTGTTTAAACTTGCC
•	1690	1710	173.0
10	TGTTCAGAGCACTCATTC	CTTCCCACCCCAGTCCTG	TCCTATCACTCTAATTCGGATTT
	1750	1770	1790
15	GCCATAGCCTTGAGGTTA	TGTCCTTTTCCATTAAGTA	CATGTGCCAGGAAACAGCGAGAG
	1810	1830	1850
20	AGAGAAAGTAAACGGCAG	TAATGCTTCTCCTATTTCT	CCAAAGCCTTGTGTGAACTAGCA
	1870	1890	1910
25	AAGAGAAGAAAATCAAAT	ATATAACCAATAGTGAAAT	GCCACAGGTTTGTCCACTGTCAG
	1930	1950	1970
	GGTTGTCTACCTGTAGGA	TCAGGGTCTAAGCACCTTG	GTGCTTAGCTAGAATACCACCTA
30	1990	2010	2030
	ATCCTTCTGGCAAGCCTG	TCTTCAGAGAACCCACTAG	AAGCAACTAGGAAAAATCACTTG
35	2050	2070	2090
	CCAAAATCCAAGGCAATT	CCTGATGGAAAATGCAAAA	GCACATATATGTTTAATATCTT
40	2110	2130	2150
45	TATGGGCTCTGTTCAAGG	CAGTGCTGAGAGGGAGGGG	TTATAGCTTCAGGAGGAACCAG
	2170	2190	2210
50	CTTCTGATAAACACAATC	TGCTAGGAACTTGGGAAAG	GAATCAGAGAGCTGCCCTTCAGC

(

	2230	2250	2270				
	GATTATTTAAATTGTTAA	NGAATACACAATTTGGGGTA	rtgggatttttctcttttctc	:			
5	2290	2310	2330				
	TGAGACATTCEAECATTT:	IAATTTTGTAACTGCTTAT	TITITATODOAAAAOTOTATIT	[			
10	2350	2370	2390				
	ACTTAGCTTAGCTATGTC	AGCCAATCCGATTGCCTTAG	GTGAAAGAAACCACCGAAATC	-			
15	2410 -	2430	2450				
	CTCAGGTCCCTTGGTCAG	GAGCCTCTCAAGATTTTTTT	TGTCAGAGGCTCCAAATAGAA	A			
20	2470	2490	2510				
	ATAAGAAAAGGTTTTCTT	CATTCATGGCTAGAGCTAGA	TTTAACTCAGTTTCTAGGCAC	C			
25	2530	2550	2570				
	TCAGACCAATCATCAACT	ACCATTCTATTCCATGTTTG	CACCTGTGCATTTTCTGTTTG	Ċ			
30	2590	2610	2630				
	CCCCATTCACTTTGTCAG	CCCCATTCACTTTGTCAGGAAACCTTGGCCTCTGCTAAGGTGTATTTGGTCCTTGAGAAG					
ne.	2650	2670	2690				
35	TGGGAGCACCCTACAGGG	ACACTATCACTCATGCTGG	rggcattgtttacagctagaaa	G			
	2710	2730	2750				
<del>1</del> 0	CTGCACTGGTGCTAATGC	CCCTTGGGAAATGGGGCTG	PGAGGAGGAGGATTATAACTTA	٠C			
	2770	2790	2810				
<b>4</b> 5	GCCTAGCCTCTTTTAACA	GCCTCTGAAATTTATCTTT	rcttctatggggtctataaatc	T			
	2830	2850	2870				
50	ATCTTATAATAAAAAGG	AGGACAGGAGGAAGACAGG	CAAATGTACTTCTCACCCAGT	T			

	2890	2910	2930
7	CTACACAGATGGAATCT	CTTTGGGGCTAAGAGAAAGGT	TTTATTCTATATTGCTTACCT
5			
	2950	2970	2990
C	SATCTCATGTTAGGCCTA	AGAGGCTTTCTCCAGGAGGAT	TTAGCTTGGAGTTCTCTATACT
10			
	3010	3030	3050
(	CAGGTACCTCTTTCAGGG	TTTTCTAACCCTGACACGGA	TGTGCATACTTTCCCTCATCC
15			
	3070	3090	3110
j	ATGCTGTGCTGTTATT	TAATTTTTCCTGGCTAAGAT	CATGTCTGAATTATGTATGAAA
20			
	3130	3150	3170
4	ATTATTCTATGTTTTTA1	AATAAAAATAATATCAGA	CATCGARAAAAAAA,
25			
30			
35			
40			
45			
50			

(2)

	430	450	470
5	ThrGlyPheTyrThrLeuGl	CAGTCATAAAGTCAGATCT lnValileLysSerAspLe	TGTGAATGAAGAAGCAACTGGA uValAsnGluGluAlaThrGly
	490	510	530
10			CTCCAGCAACAACTCCAACCCT eSerSerAsnAsnSerAsnPro
	550	570	590
15	GTGGAGGACAAGGATGCTG' ValGluAspLysAspAlaV	TGGCCTTCACCTGTGAACC alalaPheThrCysGluPr	TGAGACTCAGGACACAACCTAC oGluThrGlnAspThrThrTyr
	610	630	650
20	CTGTGGTGGATAAACAATC. LeuTrpTrpIleAsnAsnG	AGAGCCTCCCGGTCAGTCC lnSerLeuProValSerPr	CAGGCTGCAGCTGTCCAATGGC OArgLeuGlnLeuSerAsnGly
25	670	690	710
			CACAGGACCCTATGAGTGTGAA pThrGlyProTyrGluCysGlu
30	730	750	770
			CACCTTGAATGTCACCTATGGC lThrLeuAsnValThrTyrGly
35	790	810	830
40			CCGTCCAGGGGCAAACCTCAGC rArgProGlyAlaAsnLeuSer
45			
50			

	1270	1290	1310	
5			CATCATGCTGAACGTAAACTAT DIleMetLeuAsnValAsnTyr	
	1330	1350	1370	
10			CATTGCTGGCATTGTGATTGGA alleAlaGlyIleValIleGly	
	1390	1410	1430	
15	GTAGTGGCCCTGGTTGCT ValValAlaLeuValAla	CTGATAGCAGTAGCCCTGGC LeulleAlaValAlaLeuAl	ATGTTTTCTGCATTTCGGGAAG aCysPheLeuHisPheGlyLys	
20	1450	1470	1490	
20	ACCGGCAGCTCAGGACCA ThrGlySerSerGlyPro		AGATGAATGAAGTTACTTATTC	
25	1510	1530	1550	···
	TACCCTGAACTTTGAAGC	CCAGCAACCACACCAA	CTTCAGCCTCCCCATCCCTAAC	
30	1570	1590	1610	
	AGCCACAGAAATAATTTA	TTCAGAAGTAAAAAAGCAGT	AAAAAAAAAADTOOAAADTAA	
35	1630			
40				
40				
			,	
45				

(3)

	10	30	50
	CAGCCGTGCTCGAAGCGTT	CCTGGAGCCCAAGCTCTCC1	CCACAGGTGAAGACAGGGCCA
10			
	70	90	110
15			CAGAGTGCGTGTACCCTGGCAG SArgValArgValProTrpGln
	130	150	170
20			CCCGCCCACCACTGCCCAGCTC
	190	210	230
25			GGAGGTTCTTCTCCTTGTCCAC sGluValLeuLeuLeuValHis
	250	270	290
30			AGGGGAAAGAGTGGATGGCAAC sGlyGluArgValAspGlyAsn
35	310	330	350
			TACCCCAGGGCCCGCAAACAGG aThrProGlyProAlaAsnSer
40	370	390	410

45 .

(4)

5	1	acagcacagctgacagccgtactcaggaagcttctggatcctaggcttatctccacagag	60
J	61	gagaacacacagcagcagagaccatggggcccctctcagcccctccct	120
10	1.21.	atcacttggaaggggtcctgctcacagcatcacttttaaacttctggaatccgcccaca IleThrTrpLysGlyValLeuLeuThrAlaSerLeuLeuAsnPheTrpAsnProProThr	1.80
	181	actgcccaagtcacgattgaagcccagccacccaaagtttctgaggggaaggatgttctt ThrAlaGlnValThrIleGluAlaGlnProProLysValSerGluGlyLysAspValLeu	240
15	241	ctacttgtccacaatttgccccagaatcttgctggctacatttggtacaaagggcaaatg LeuLeuValHisAsnLeuProGlnAsnLeuAlaGlyTyrIleTrpTyrLysGlyGlnMet	300
15	301	acatacgtctaccattacattacatcatatgtagtagacgqtcaaagaattatatatggg ThrTyrValTyrHisTyrIleThrSerTyrValValAspGlyGlnArgIleIleTyrGly	360
	361	cctgcatacagtggaagagaaagagtatattccaatgcatccctgctgatccagaatgtc ProAlaTyrSerGlyArgGluArgValTyrSerAsnAlaSerLeuLeuIleGlnAsnVal	420
20	.421.	acgcaggaggatgcaggatcctacaccttacacatcataaagcgacgcgatgggactgga ThrGlnGluAspAlaGlySerTyrThrLeuHisIleIleLysArgArgAspGlyThrGly	480
	481	ggagtaactggacatttcaccttcaccttacacctggagactcccaagccctccatctcc GlyValThrGlyHisPheThrPheThrLeuHisLeuGluThrProLysProSerIleSer	540
25	541	agcagcaacttaaatcccagggaggccatggaggctgtgatcttaacctgtgatcctgcg SerSerAsnLeuAsnProArgGluAlaMetGluAlaValIleLeuThrCysAspProAla	600
	601	actccagccgcaagctaccagtggtggatgaatggtcagagcctccctatgactcacagg ThrProAlaAlaSerTyrGlnTrpTrpMetAsnGlyGlnSerLeuProMetThrHisArg	660
30	661	ttgcagctgtccaaaaccaacaggaccctctttatattttggtgtcacaaagtatattgca LeuGlnLeuSerLysThrAsnArgThrLeuPheIlePheGlyValThrLysTyrIleAla	720
	721	ggaccctatgaatgtgaaatacggaacccagtgagtgccagccgcagtgacccagtcacc GlyProTyrGluCysGluIleArgAsnProValSerAlaSerArgSerAspProValThr	780
35	781	ctgaatctcctcccaaagctgtccaagccctacatcacaatcaacaacttaaaccccaga LeuAsnLeuLeuProLysLeuSerLysProTyrIleThrIleAsnAsnLeuAsnProArg	840
	841	gagaataaggatgtcttaaccttcacctgtgaacctaagagtgagaactacacctacatt GluAsnLysAspValLeuThrPheThrCysGluProLysSerGluAsnTyrThrTyrIle	900
40	901	tggtggctaaatggtcagagcctccctgtcagtcccagggtaaagcgacccattgaaaac TrpTrpLeuAsnGlyGlnSerLeuProValSerProArgValLysArgProIleGluAsn	960
	961	aggatcctcattctacccaatgtcacgagaaatgaaacaggaccttatcaatgtgaaata ArgileLeuileLeuProAsnValThrArgAsnGluThrGlyProTyrGlnCysGluile	1020
45	1021	cgggaccgatatggtggcatccgcagtgacccagtcaccctgaatgtcctctatggtcca ArgAspArgTyrGlyGlyIleArgSerAspProValThrLeuAspValLeuTvrGlyPro	1080

	1081	gacctccccagcatttacccttcattcacctattaccgttcaggagaaaacctctacttt AspLeuProSerIleTyrProSerPheThrTyrTyrArgSerGlyGluAsnLeuTyrPhe	1140
5	1141	tcctgcttcggtgagtctaacccacgggcacaatattcttggacaattaatgggaagttt SerCysPheGlyGluSerAsnProArgAlaGlnTyrSerTrpThrIleAsnGlyLysPhe	1200
-	1201 -	cagctatcaggacaaaagctctctatccccaaataactacaaagcatagtgggctctat GlnLeuSerGlyGlnLysLeuSerIleProGlnIleThrThrLysHisSerGlyLeuTyr	1260
10	1261	gcttgctctgttcgtaactcagccactggcaaggaaagctccaaatccatcacagtcaaa AlaCysSerValArgAsnSerAlaThrGlyLysGluSerSerLysSerIleThrValLys	1320
	1321	gtctctgactggatattaccctgaattctactagttcctccaattccattttctcccatg ValSerAspTrpIleLeuProEnd	1380
15	1381 1441 1501	gaatcacgaagagcaagacccactctgttccagaagccctataatctggaggtggacaac tcgatgtaaatttcatgggaaaacccttgtacctgacatgtgagccactcagaactcacc aaaatgttcgacaccataacaacagctactcaaactgtaaaccaggataagaagttgatg	1440 1500 1560
	1561 1621 1681	acttcacactgtggacagtttttcaaagatgtcataacaagactccccatcatgacaagg ctccaccctctactgtctgctcatgcctgcctctttcacttggcaggataatgcagtcat tagaatttcacatgtagtagcttctgagggtaacaacagagtgtcagatatgtcatctca	1620 1680 1740
20	1741 1801 1861 1921 1981	acctcaaacttttacgtaacatctcagggaaatgtggctctctccatcttgcatacaggg ctcccaatagaaatgaacacagagatattgcctgtgtgttttgcagagaagatggtttcta taaagagtaggaaagctgaaattatagtagagtctcctttaaatgcacattgtgtggatg gctctcaccatttcctaagagatacagtgtaaaaacgtgacagtaatactgattctagca gaataaacatgtaccacatttgcaaaaaa	1800 1860 1920 1980 2010

25 and

(5)

5	1	gggtggatcctaggctcatctccataggggagaacacacatacagcagagaccatggga MetGly	59
	60	cccctctcagcccctccctgcactcagcacatcacctggaaggggctcctgctcacagca ProLeuSerAlaProProCysThrGlnHisIleThrTrpLysGlyLeuLeuLeuThrAla	119
10	120	tcacttttaaacttctggaacctgcccaccactgcccaagtaataattgaagcccagcca SerLeuLeuAsnPheTrpAsnLeuProThrThrAlaGlnVallleIleGluAlaGlnPro	179
	180	cccaaagtttctgaggggaaggatgttcttctacttgtccacaatttgccccagaatctt ProLysValSerGluGlyLysAspValLeuLeuLeuValHisAsnLeuProGlnAsnLeu	239
15	240	actggctacatctggtacaaagggcaaatgacggacctctaccattacattacatcatat ThrGlyTyrIleTrpTyrLysGlyGlnMetThrAspLeuTyrHisTyrIleThrSerTyr	299
•	<u>3</u> 00	gtagtagacggtcaaattatatatgggcctgcctacagtggacgagaaacagtatattcc ValValAspGlyGlnIleIleTyrGlyProAlaTyrSerGlyArgGluThrValTyrSer	359
20	360	aatgcatccctgctgatccagaatgtcacacaggaggatgcaggatcctacaccttacac AsnAlaSerLeuLeuIleGlnAsnValThrGlnGluAspAlaGlySerTyrThrLeuHis	419
	420	atcataaagcgaggcgatgggactggaggagtaactggatatttcactgtcaccttatac IleIleLysArgGlyAspGlyThrGlyGlyValThrGlyTyrPheThrValThrLeuTyr	479
25	480	tcggagactcccaagcgctccatctccagcagcaacttaaaccccagggaggtcatggag SerGluThrProLysArgSerIleSerSerSerAsnLeuAsnProArgGluValMetGlu	539
	540	gctgtgcgcttaatctgtgatcctgagactccggatgcaagctacctgtggttgctgaat AlaValArgLeuIleCysAspProGluThrProAspAlaSerTyrLeuTrpLeuLeuAsn	599
30	600	ggtcagaacctccctatgactcacaggttgcagctgtccaaaaccaacaggaccctctat GlyGlnAsnLeuProMetThrHisArgLeuGlnLeuSerLysThrAsnArgThrLeuTyr	659
	660 _)	ctatttggtgtcacaaagtatattgcagggccctatgaatgtgaaatacggaggggagtg LeuPheGlyValThrLysTyrIleAlaGlyProTyrGluCysGluIleArgArgGlyVal	719
35	720	agtgccagccgcagtgacccagtcaccctgaatctcctcccgaagctgcccatgccttac SerAlaSerArgSerAspProValThrLeuAsnLeuLeuProLysLeuProMetProTyr	779
	780	atcaccatcaacaacttaaaccccagggagaagaaggatgtgttagccttcacctgtgaa IleThrIleAsnAsnLeuAsnProArgGluLysLysAspValLeuAlaPheThrCysGlu	839
40	840	cctaagagtcggaactacacctacatttggtggctaaatggtcagagcctcccggtcagt ProLysSerArgAsnTyrThrTyrIleTrpTrpLeuAsnGlyGlnSerLeuProValSer	899
	900	ccgagggtaaagcgacccattgaaaacaggatactcattctacccagtgtcacgagaaat ProArgValLysArgProIleGluAsnArgIleLeuIleLeuProSerValThrArgAsn	959
45	960	gaaacaggaccctatcaatgtgaaatacgggaccgatatggtggcatccgcagtaaccca GluThrGlyProTyrGlnCysGluIleArgAspArgTyrGlyGlyIleArgSerAsnPro	1019

	1020	gtcaccctgaatgtcctctatggtccagacctccccagaatttacccttacttcacctat ValThrLeuAsnValLeuTyrGlyProAspLeuProArgIleTyrProTyrPheThrTyr	1079
5	1080	taccgttcaggagaaaacctcgacttgtcctgctttgcggactctaacccaccggcagag TyrArgSerGlyGluAsnLeuAspLeuSerCysPheAlaAspSerAsnProProAlaGlu	1139
	1140	tatttttggacaattaatgggaagtttcagctatcaggacaaaagctctttatcccccaa TyrPheTrpThrIleAsnGlyLysPheGlnLeuSerGlyGlnLysLeuPheIleProGln	1199
10	1200	attactacaaatcatagcgggctctatgcttgctctgttcgtaactcagccactggcaag IleThrThrAsnHisSerGlyLeuTyrAlaCysSerValArgAsnSerAlaThrGlyLys	1259
	1260	gaaatctccaaatccatgatagtcaaagtctctggtccctgccatggaaaccagacaga	1319
15	1320	tctcattaatggctgccacaatagagacactgagaaaaagaacaggttgataccttcatg SerHisEnd	1379
	1380 1440 1500 1560	aaattcaagacaaagaagaaaaaggctcaatgttattggactaaataatcaaaaggataa tgttttcataatttttattggaaaatgtgctgattcttggaatgttttattctccagatt tatgaacttttttttcttcagcaattggtaaagtatacttttgtaaacaaaaattgaaaca tttgcttttgctctctatctgagtgcccccc 1591	1439 1499 1559
20			

- 2. A replicable recombinant cloning vehicle having an insert comprising a nucleic acid of claim 1.
- 25 3. A cell that is transfected, infected or injected with a recombinant cloning vehicle of claim 2.
  - 4. A method for preparing a polypeptide, said method comprising the steps of
    - (a) culturing the cell of claim 3
    - (b) recovering the polypeptide expressed by said cell.
  - 5. A method for preparing an antibody directed against a polypeptide said method comprising the steps of (a) preparing said polypeptide by the method of claim 4
    - (b) injecting said polypeptide into a host capable of producing antibodies and
    - (c) recovering said antibodies.

### Patentansprüche

1. Nucleinsäure, umfassend eine Basen-Sequenz, die für eine Peptid-Sequenz codiert, dadurch gekennzeichnet, daß die Gruppen-Nucleinsäure eine DNA ist, die aus der folgenden Gruppe von fünf Sequenzen ausgewählt ist:

30

35

40

45

	10	30	50
	CAGCCGTGCTCGAAGCGTTC	CCTGGAGCCCAAGCTCTCCT	CCACAGGTGAAGACAGGGCCA
5			
	7 0	90	110
10			AGAGTGCGTGTACCCTGCCAG ArgValArgValProTrpGln
			•
	130	150	170
15			CCGCCCACCACTGCCCAGCTC PROPROTHITHIALAGINESS
	190	210	230
20	The The Gluber met Prop	PheAsnValAlaGluGlyLy	GGAGGTTCTTCTCCTTGTCCAI sGluValLeuLeuLeuValHis
	250	270	290
25			AGGGGAAAGAGTGGATGGCAAC SClyGluArgValAspGlyAsn
	310	330	350 -
30			TACCCCAGGGCCCGCAAACAGC aThrProGlyProAlaAsnSer
	370	390	410
35			CCAGAACGTCACCCAGAATGAC eGlnAsnVslThrGlnAsnAsp
	130	450	470
40	ACAGGATTCTACACCCTA ThrGlyPheTyrThrLeu	CAAGTCATAAAGTCAGATCT GlnVallleLysSerAspLe	TTGTGAATGAAGAAGCAACTGGA EUValAsnGluGluAlaThrGly
45			

	1030	1050	1070		
			CTCACCTGGGGCCATTGCTGGC uSerProGlyAlaIleAlaGly		
5					
	1093	1110	1130		
10			AGTAGCCCTGGCATGTTTTCTC aValAlaLeuAlaCysPheLeu		
	1150	1170	1190		
15			ATCTCACAGAGCACAAACCCTCA spLeuThrGluHisLysProSer		
	1210	1230	1250		
20			CTAACAAGATGAATGAAGTTACT roAsnLysMetAsnGluValTh:		
	1270	1290	1310		
25			RACCAACTTCAGCCTCCCCATCC InProThrSerAlaSerProSer		
	1330	1350	1370		
30	CTAACAGCCACAGAAATAATTTATTCAGAAGTAAAAAGCAGTAATGAAACCTGTCCTGC LeuThralaThrGluIleIleTyrSerGluValLysLysGln				
35	1390	1410	1430		
	TCACTGCAGTGCTGATG	PATTTCAAGTCTCTCACCCT	CATCACTAGGAGATTCCTTTCCC		
40	1450	1470	1490		
	CTGTAGGGTAGAGGGGT	GGGGACAGAACAACTTTCT	CCTACTCTTCCTTCCTAATAGGC		
45	1510	1530	1550		
ATCTCCAGGCTGCCTGGTCACTGCCCCTCTCAGTGTCAATAGATGAAAGT					
	1570	1590	1610		
50	AGTCTGTAGGAAACCCA	ACCTTCTTGTCATTGAAAT	PTGGCAAAGCTGACTTTGGGAAAG		

	1630	1650	1670
5	AGGGACCAGAACTTCCCCT	CCCTTCCCCCAAC	CTGGACTTGTTTTAAACTTGCC
	1690	1710	1730 .
	TGTTCAGAGCACTCATTCC	TTCCCACCCCCAGTCCTGT	CCTATCACTCTAATTCGGATTT
10	1750	1770	1790
	GCCATAGCCTTGAGGTTAT	GTCCTTTTCCATTAAGTAC	DAGGODDAAACAGCGAGAG
15	1810	1830	1850
	AGAGAAAGTAAACGGCAGT	AATGCTTCTCCTATTTCTC	CAAAGCCTTGTGTGAACTAGCA
20	1870	1890 -	1910
	AAGAGAAGAAAATCAAATA	TATAACCAATAGTGAAATG	CCACAGGTTTGTCCACTGTCAG
25	1930	1950	1970
	GGTTGTCTACCTGTAGGAT	CAGGGTCTAAGCACCTTGC	TGCTTAGCTAGAATACCACCTA
30	1990	2010	2030
	ATCCTTCTGGCAAGCCTGT	CCTTCAGAGAACCCACTAGA	AGCAACTAGGAAAAATCACTTO
35	2050	2070	2090
	CCAAAATCCAAGGCAATTO	CTGATGGARAATGCAARA	GCACATATATGTTTAATATCTT
40	2110	2130	2150
	TATGGGCTCTGTTCAAGG	CAGTGCTGAGAGGGAGGGG	PTATAGCTTCAGGAGGGAACCAC
45	2170	2190	. 2210
	•		SAATCAGAGAGCTGCCCTTCAG

	2230	2250	2270	
	GATTATTTAAATTGTTAAA	GAATACACAATTTGGGGTA	ITGGGATTTTTCTCCITTTCTC	
5				
	2295	2310	2330	
	TGAGACATTCCACCATTT	AATTITTGTAACTGCTTAT	TTTTTADDDA44ADTDTATT	
10				
	2350	2370	2390	
	ACTTAGCTTAGCTATGTC	GCCAATCCGATTGCCTTAG	GTGAAAGAAACCNCCGAAATC	•
15	2410 .	2430	` 2450	
	•		•	,
	CICAGGICCCIIGGICAGG	SAGCCICICAAGAIIIITT	TGTCAGAGGCTCCAAATAGAA	
20	2470	2490	2510	
	ATAAGAAAAGGTTTTCTT	CATTCATGGCTAGAGCTAGA	TTTAACTCAGTTTCTAGGCAC	:
25	2530	2550	2570	
	TCAGACCAATCATCAACT	ACCATTCTATTCCATGTTTG	CACCTGTGCATTTTCTGTTTG	:
30	2590	2610	2630	
	CCCCATTCACTTTGTCAG	GAAACCTTGGCCTCTGCTAA	GGTGTATTTGGTCCTTGAGAA	G
35	2650	2670	2 5 9 C	
	TGGGAGCACCCTACAGGG	ACACTATCACTCATGCTGGT	PGGCATTGTTTACAGCTAGA&&	G
40	2710	2730	2750	
	CTGCACTGGTGCTAATGC	CCCTTGGGAAATGGGGCTG	ATTO ANTATTADO ADDADO ADT	Ġ
45	2770	2790	2810	
	GCCTAGCCTCTTTTAACA	GCCTCTGAAATTTATCTTT	TCTTCTATGGGGTCTATAAATC	. T
	2830	2850	2870	
50	ATCTTATA			

# This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record.

# **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS
IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
FADED TEXT OR DRAWING
BLURRED OR ILLEGIBLE TEXT OR DRAWING
☐ SKEWED/SLANTED IMAGES
☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
☐ GRAY SCALE DOCUMENTS
☐ LINES OR MARKS ON ORIGINAL DOCUMENT
☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
□ other:

# IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.